GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 1999, 10:12:06 ; Search time 24.98 Seconds
(without alignments)
222.657 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLLPEDGDRNA.....INKIYQPGESKTVEQALGLI 275

Scoring table: BLOSUM62

Searched: 162890 seqs, 20225328 residues

Database : A_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	95.5	6.5	531	1	R48994	Human glycosyltran
2	95.5	6.5	531	1	R97615	Rat N-acetylglucos
3	93.5	6.4	531	1	W24015	Human N-acetylgluc
4	90	6.2	1780	1	W53863	Human gravin polyp
5	88.5	6.1	1684	1	W25671	hABC3 protein. New
6	88.5	6.1	1684	1	W46761	Amino acid sequenc
7	88.5	6.1	1704	1	W46771	Amino acid sequenc
8	87.5	6.0	536	1	R97614	Rat N-acetylglucos
9	87.5	6.0	536	1	W24014	Rat N-acetylglucos
10	87	6.0	1257	1	R87628	Alpha-ketoglutaric
11	87	6.0	1257	1	W41781	B. lactofermentum
12	86.5	5.9	675	1	R29571	Human heregulin-be
13	86.5	5.9	675	1	W74487	Amino acid sequenc
14	86.5	5.9	675	1	W74488	Amino acid sequenc
15	86.5	5.9	675	1	W74486	Amino acid sequenc
16	86.5	5.9	675	1	W74485	Degenerate amino a
17	86.5	5.9	675	1	W74504	Amino acid sequenc
18	86.5	5.9	675	1	W74503	Amino acid sequenc
19	86.5	5.9	675	1	W74502	Amino acid sequenc
20	86.5	5.9	675	1	W74501	Amino acid sequenc
21	86.5	5.9	675	1	W74500	Amino acid sequenc
22	86.5	5.9	675	1	W74499	Amino acid sequenc
23	86.5	5.9	675	1	W74498	Amino acid sequenc
24	86.5	5.9	675	1	W74497	Amino acid sequenc
25	86.5	5.9	675	1	W74496	Amino acid sequenc
26	86.5	5.9	675	1	W74495	Amino acid sequenc
27	86.5	5.9	675	1	W74494	Amino acid sequenc
28	86.5	5.9	675	1	W74493	Amino acid sequenc
29	86.5	5.9	675	1	W74492	Amino acid sequenc
30	86.5	5.9	675	1	W74491	Amino acid sequenc
31	86.5	5.9	675	1	W74490	Amino acid sequenc
32	86.5	5.9	675	1	W74489	Amino acid sequenc
33	86.5	5.9	675	1	W74509	Amino acid sequenc
34	86.5	5.9	675	1	W74508	Amino acid sequenc
35	86.5	5.9	675	1	W74507	Amino acid sequenc
36	86.5	5.9	675	1	W74506	Amino acid sequenc
37	86.5	5.9	675	1	W74505	Amino acid sequenc
38	85.5	5.8	669	1	R29570	Human heregulin-al
39	85.5	5.8	551	1	R68565	Human NDF-betala c
40	82.5	5.6	637	1	R29573	Human heregulin-be
41	82.5	5.6	849	1	W48102	Human discs large
42	82.5	5.6	733	1	W60790	Deletion mutant de
43	80.5	5.5	2165	1	R94930	RSV RNA-dependent
44	80.5	5.5	2165	1	W77015	Respiratory syncyt
45	79	5.4	621	1	W76213	Human ELL protein.

ALIGNMENTS

RESULT 1
R48994
ID R48994 standard; Protein; 531 AA.
AC R48994;
DT 04-SEP-1994 (first entry)
DE Human glycosyltransferase GnT-III protein.
KW Glycosyltransferase; enzyme; cancer diagnosis; ss.
OS Homo sapiens.
PN EP-585083-A.
PD 02-MAR-1994.
PF 20-AUG-1993; 306628.
PR 21-AUG-1992; JP-243984.
PA (TAKI) TAKARA SHUZO CO LTD.
PI Nishikawa A, Taniguchi N;
DR WPI; 94-067563/09.
DR N-PSDB; Q57973.
PT New gene for human glycosyltransferase GnT-III - and related
PT vectors and transformed cells, useful in diagnosis of cancer
PS Claim 3; Page 10; 14pp; English.
CC This glycosyltransferase protein is human UDP-N-acetylglucosamine;
CC beta-D-mannoside-beta1-4N-acetyl-glucosaminyltransferase. It may
CC be expressed recombinantly in host cells and used in cancer diagnosis.
SQ Sequence 531 AA;

Query Match 6.5%; Score 95.5; DB 1; Length 531;
Best Local Similarity 22.6%; Pred. No. 0.088;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;

```
Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || |  :  :  || |  ||  ||  ||:    |   |  :||     | | :
Db      53 PVTPQASPEPGGPDLLRT-PLYSHSPLLQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
             |             ||  :  |:  | |     |||||: |   : :|    ||  :
Db     111 AG-------------GVCFKPGTKMLERP-----PPGRPEEKPEGANGSSARRPPRYLL 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
           :              |   ||         | ||:     |               : |:
Db     152 SAR----------ERTGGR-----GARRKWVECVCLPGW-------------HGPSC 180

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
           |  :   :   :|  :: :|    || |:        |||
Db     181 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVC 236

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
                 | :||   :   | |   |||||: | :           : || |    ||
Db     237 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 292
```

RESULT 2
R97615
ID R97615 standard; Protein; 531 AA.
AC R97615;
DT 21-NOV-1996 (first entry)
DE Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.
KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;
KW neoplasia; invasive.
OS Rattus sp.
PN J08109139-A.
PD 30-APR-1996.
PF 12-OCT-1994; 271802.
PR 12-OCT-1994; JP-271802.
PA (TAKI) TAKARA SHUZO CO LTD.
DR WPI; 96-263788/27.
PT Inhibitor of cancer metastasis - contains N-acetyl:glucosaminyl

PT transferase-III isolated from rat kidneys
PS Claim 3; Page 8-10; 12pp; Japanese.
CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl
CC transferase-III (GnT-III). These clones are useful to reinforce
CC the activity of GnT-III present in cancer cells and cells surrounding
CC cancer cells for the inhibition of metastasis, preventing the spread
CC of cancer cells to tissues other than that of their origin. The
CC clones were isolated from a rat kidney homogenate.
SQ Sequence 531 AA;

Query Match 6.5%; Score 95.5; DB 1; Length 531;
Best Local Similarity 22.6%; Pred. No. 0.088;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;

```
Qy      6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
          |:     || |  :  :  || |  ||  ||  ||:    |  | :||     | | :
Db     53 PVTPQASPEPGGPDLLRT-PLYSHSPLLQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 110

Qy     64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
          |            ||  :  |:  | |     |||||  : |   : :|    ||  :
Db    111 AG-------------GVCFKPGTKMLERP-----PPGRPEEKPEGANGSSARRPPRYLL 151

Qy    124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
          :              |   ||        | ||:     |               : |:
Db    152 SAR-----------ERTGGR-----GARRKWVECVCLPGW-------------HGPSC 180

Qy    179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
          |  :   :   :|  :: :|    || |:        |||
Db    181 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVC 236

Qy    213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
              |  :||   :   | |   |||||  | :           || |    ||
Db    237 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 292
```

RESULT 3
W24015
ID W24015 standard; Protein; 531 AA.
AC W24015;
DT 10-FEB-1998 (first entry)
DE Human N-acetylglucosaminyl transferase III.
KW N-acetylglucosaminyl transferase III; GnT-III; rat; human;
KW virus; replication; inhibitor; hepatitis B; hepatitis C; HIV;
KW viral disease; human immunodeficiency virus.
OS Homo sapiens.
PN WO9718836-A1.
PD 29-MAY-1997.
PF 17-JUL-1996; J01986.
PR 17-NOV-1995; JP-322474.
PA (TAKI) TAKARA SHUZO CO LTD.
PI Ihara Y, Miyoshi E, Taniguchi N;
DR WPI; 97-297877/27.
DR N-PSDB; T77094.
PT Virus replication inhibitor containing N-acetyl:glucosaminyl
PT transferase III or its gene - useful in treatment of hepatitis B, C
PT and HIV viral infections
PS Claim 6; Page 22-26; 12pp; Japanese.
CC The present sequence represents human N-acetylglucosaminyl transferase
CC III (GnT-III). The present specification describes a virus replication
CC inhibitor which contains the gene encoding GnT-III as an active
CC ingredient. The virus replication inhibitor is used against hepatitis B,
CC C and HIV. It is used for the treatment of viral diseases.
SQ Sequence 531 AA;

Query Match 6.4%; Score 93.5; DB 1; Length 531;
Best Local Similarity 22.6%; Pred. No. 0.14;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;

```
Qy      6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
          |:     || |  :  :  || |  ||  ||  ||:    |  | :||     | | :
Db     53 PVTPQASPEPGGPDLLRT-PLYSHSPLLQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 110

Qy     64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
          |            ||  :  |:  | |     |||||  : |   : :|    ||  :
Db    111 PG-------------GVCFKPGTKMLERP-----PPGRPEEKPEGANGSSARRPPRYLL 151

Qy    124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
          :              |   ||        | ||:     |               : |:
Db    152 SAR-----------ERTGGR-----GARRKWVECVCLPGW-------------HGPSC 180

Qy    179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
          |  :   :   :|  :: :|    || |:        |||
Db    181 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVC 236

Qy    213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
              |  :||   :   | |   |||||  | :           || |    ||
Db    237 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 292
```

RESULT 4
W53863
ID W53863 standard; peptide; 1780 AA.
AC W53863;
DT 13-JUL-1998 (first entry)
DE Human gravin polypeptide.
KW Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW Myasthenia gravis; nicotinic acetylcholine receptor.
OS Homo sapiens.
PN US5741890-A.
PD 21-APR-1998.
PF 19-DEC-1996; 769309.
PR 19-DEC-1996; US-769309.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Klauck TM, Nauert JB, Scott JD;
DR WPI; 98-260552/23.
DR N-PSDB; V23545.
PT New polypeptide fragments of protein kinase binding protein gravin -
PT are useful for the study of modulation of action between gravin and
PT protein kinase(s)
PS Example 1; Column 19-32; 32pp; English.
CC This sequence corresponds to the human gravin polypeptide, and represents
CC a polypeptide of the invention. The polypeptides are fragments capable of
CC binding to type II regulatory subunit of cAMP-dependent protein kinase
CC (PKA). Gravin is a kinase anchoring protein that binds to type II
CC regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC develops antibodies against their own nicotinic acetylcholine receptors.
CC The polypeptides are useful for providing analogues of gravin in the
CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC interactions between gravin and kinase. The peptides are involved in the
CC modulation of gravin-kinase interactions.
SQ Sequence 1780 AA;

Query Match 6.2%; Score 90; DB 1; Length 1780;
Best Local Similarity 22.6%; Pred. No. 1.8;
Matches 53; Conservative 32; Mismatches 93; Indels 56; Gaps 10;

```
Qy     31 PLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLGNLLTTIKHHPSEIIGVLPEDYTRAD 90
          | | | ||: : :|::  :      | ||:|     :  :  |  : |    |: :||
Db    663 PKP-EEPKRKVDTSVSWEALICVGSSKKRARRRSS---SDEEGGPKAMGG----DHQKAD 714

Qy     91 EE----------PGRQGRPPGRPRKMPRHESS------TSLMESPRKTMTRDSKIMFE 132
          |           | |   ||:    |   |        |  || :: :|   |   :
Db    715 EAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSK 774

Qy    133 LRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPIKAQRTDYAPNLAVDYLACREI---H 189
          |  |  : |||  ||           |   : || | :           ::: ::
Db    775 LEEKSEDSIAGSGVE----------HSTPDTEPGKEE----------SWVSIKKFIPGR 813

Qy    190 RMPRPDKSIPELPIV---PSRIDEFDATVD---PRYETDLKNEYIRHWKQVKKG 237
          |  |||     : |:     |:  :| |: |     |  | |          :| :||
Db    814 RKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKG 867
```

```
RESULT   5
W25671
ID   W25671 standard; Protein; 1684 AA.
AC   W25671;
DT   12-NOV-1997  (first entry)
DE   hABC3 protein.
KW   Human; netrin; ATPase binding cassette transporter; ribosomal L3;
KW   augmenter of liver regeneration; hNET; hABC3; SEM L3; hALR;
KW   chromosome 16; exon trapping; axon; chicken; laminin domain;
KW   C. elegans; UNC-6; cystic fibrosis.
OS   Homo sapiens.
PN   WO9702346-A2.
PD   23-JAN-1997.
PF   17-JUN-1996; U10469.
PR   30-JUN-1995; US-000596.
PA   (GENZ ) GENZYME CORP.
PI   Burn TC,  Connors TD,  Dackowski WR,  Klinger KW,  Landes GM;
PI   Van RAAY TJ;
DR   WPI; 97-108959/10.
DR   N-PSDB; T85474.
PT   New isolated human chromosome 16 genes - encode netrin, ATPase
PT   binding cassette transporter, ribosomal L3 sub-type or augmenter of
PT   liver regeneration
PS   Claim 40; Fig 8; 98pp; English.
CC   This sequence represents the ATPase binding cassette transporter protein
CC   (hABC3).  The genomic sequence was isolated from human chromosome 16
CC   by exon trapping.  hABC3 cDNA contains an open reading frame of
CC   1685 amino acids.  Comparison of ABC1, ABC2 and hABC3 reveals
CC   significant conservation in the regions surrounding the two ATP binding
CC   cassettes.  The ATP binding cassettes of hABC3 flank a large linker
CC   domain containing numerous polar residues.  The presence of these
CC   features in the linker domain suggests that this domain may play a
CC   regulatory role similar to the R domain of CFTR.  The hABC3 gene may
CC   be used in therapeutic applications for cystic fibrosis.
SQ   Sequence   1684 AA;

  Query Match             6.1%;  Score 88.5;  DB 1;  Length 1684;
  Best Local Similarity   22.0%;  Pred. No. 2.4;
  Matches   56;  Conservative   30;  Mismatches   84;  Indels   85;  Gaps   12;

Qy      39 RKLTSLLSYDPTVPESPDMKFARKRL-----GNLLTTIKHHPSEIIGVLPEDYTRADEEP 93
           | || | :  | :||  |:   | |:      :||     |   ||  | : |
Db    1323 RTLTELYTRMPVLPEDQDVADERTRILAPSPDSLL-----HTPLIIKELSKVY------- 1371

Qy      94 GRQGRPPGRPRKMPRHESSTSLMESPRKTMTRDSKIMFELRG-------KPFEMIAGRFE 146
                      |    |:   | ::       | | |          |:|: |  |
Db    1371 ---------------EQRVPLLAVDRLSLAVQKGECFGLLGFNGAGKTTTFKMLTG--E 1412

Qy     147 EEYSLGRAWVKGH-MNNEYEPIKAQRTDYAP--NLAVDYLACREIHRMPRPDKSIPELPI 203
           |  : | |:| || ::::   :: ||  | |  :  :|::  ||:  |    : |||
Db    1413 ESLTSGDAFVGGHRISSDVGKVR-QRIGYCPQFDALLDHMTGREMLVMYARLRGIPE--- 1469

Qy     204 VPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRR---TAPHARSIALINKIY 260
                              ||      | |     |     |||  :    : |
Db    1469 -------------------------RH-----IGACVENTLRGLLLEPHANKLV---RTY 1495

Qy     261 QPGESKTVEQALGLI 275
           |  : :   : ||
Db    1496 SGGNKRKLSTGIALI 1510


RESULT   6
W46761
ID   W46761 standard; Protein; 1684 AA.
AC   W46761;
DT   03-JUN-1998  (first entry)
DE   Amino acid sequence of human ATP binding cassette transporter.
KW   Human; netrin; hNET; ATP binding cassette transporter; hABC3;
KW   ribosomal L3; RPL3L; augmenter of liver regeneration; hALR; treatment;
KW   trapping; modulation; expression; antibody; identification; binding;
KW   cystic fibrosis; transport; substrate specificity; ligand; exon trap.
OS   Homo sapiens.
PN   WO9748797-A1.
PD   24-DEC-1997.
PF   16-JAN-1997; U00785.
PR   09-DEC-1996; US-762500.
PR   17-JUN-1996; US-665259.
PR   01-OCT-1996; US-720614.
PA   (GENZ ) GENZYME CORP.
PI   Burn TC,  Connors TD, Dackowski WR, Klinger KW, Landes GM,
PI   Van Raay TJ;
DR   WPI; 98-063138/06.
DR   N-PSDB; V16307.
PT   Human chromosome 16 genes encoding netrin, ATP binding cassette
PT   transporter, ribosomal L3 and augmenter of liver regeneration
PT   proteins - useful for, e.g. treatment of liver disease and cystic
PT   fibrosis
PS   Claim 47; Fig 8A-H; 220pp; English.
CC   The present sequence represents human ATP binding cassette transporter
CC   (ABC). The ABC gene is located in the PKD1 locus, between the LCN1 and
CC   D16s291 markers in a centromeric to telomeric orientation. The sequence
CC   shows homology with murine ABC1 and ABC2 genes. The ABC proteins are
CC   responsible for the transport of a wide variety of substrates across
CC   cell membranes. Proteins in this family are linked by strong structural
CC   similarities. ABC transporters govern unidirectional transport of
CC   molecules into or out of cells and across subcellular membranes. The
CC   sequence was isolated using an exon trap. Sequences encoding human netrin
CC   (hNET), human ribosomal L3 (RPL3L), and human augmenter of liver
CC   regeneration (hALR) were also isolated. The antisense oligonucleotides of
CC   the DNA sequence are used to modulate expression of ABC prevent its
CC   translation. Antibodies against ABC can be used to block binding of its
CC   naturally occurring ligands. Host cells containing vectors with DNA
CC   inserts encoding the protein can be used in a method for identifying
CC   compounds which bind to ABC. Modulation or alteration of hABC3 substrate
CC   specificity may have significant therapeutic implications for cystic
CC   fibrosis.
SQ   Sequence   1684 AA;

  Query Match             6.1%;  Score 88.5;  DB 1;  Length 1684;
  Best Local Similarity   22.0%;  Pred. No. 2.4;
  Matches   56;  Conservative   30;  Mismatches   84;  Indels   85;  Gaps   12;

Qy      39 RKLTSLLSYDPTVPESPDMKFARKRL-----GNLLTTIKHHPSEIIGVLPEDYTRADEEP 93
           | || | :  | :||  |:   | |:      :||     |   ||  | : |
Db    1323 RTLTELYTRMPVLPEDQDVADERTRILAPSPDSLL-----HTPLIIKELSKVY------- 1371

Qy      94 GRQGRPPGRPRKMPRHESSTSLMESPRKTMTRDSKIMFELRG-------KPFEMIAGRFE 146
                      |    |:   | ::       | | |          |:|: |  |
Db    1371 ---------------EQRVPLLAVDRLSLAVQKGECFGLLGFNGAGKTTTFKMLTG--E 1412

Qy     147 EEYSLGRAWVKGH-MNNEYEPIKAQRTDYAP--NLAVDYLACREIHRMPRPDKSIPELPI 203
           |  : | |:| || ::::   :: ||  | |  :  :|::  ||:  |    : |||
Db    1413 ESLTSGDAFVGGHRISSDVGKVR-QRIGYCPQFDALLDHMTGREMLVMYARLRGIPE--- 1469

Qy     204 VPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRR---TAPHARSIALINKIY 260
                              ||      | |     |     |||  :    : |
Db    1469 -------------------------RH-----IGACVENTLRGLLLEPHANKLV---RTY 1495

Qy     261 QPGESKTVEQALGLI 275
           |  : :   : ||
Db    1496 SGGNKRKLSTGIALI 1510


RESULT   7
W46771
ID   W46771 standard; Protein; 1704 AA.
AC   W46771;
DT   03-JUN-1998  (first entry)
DE   Amino acid sequence of human ATP binding cassette transporter.
KW   Human; netrin; hNET; ATP binding cassette transporter; hABC3;
KW   ribosomal L3; RPL3L; augmenter of liver regeneration; hALR; treatment;
KW   trapping; modulation; expression; antibody; identification; binding;
```

KW cystic fibrosis; transport; substrate specificity; ligand; exon trap.
OS Homo sapiens.
PN WO9748797-A1.
PD 24-DEC-1997.
PF 16-JAN-1997; U00785.
PR 09-DEC-1996; US-762500.
PR 17-JUN-1996; US-665259.
PR 01-OCT-1996; US-720614.
PA (GENZ) GENZYME CORP.
PI Burn TC, Connors TD, Dackowski WR, Klinger KW, Landes GM,
PI Van Raay TJ;
DR WPI; 98-063138/06.
PT Human chromosome 16 genes encoding netrin, ATP binding cassette
PT transporter, ribosomal L3 and augmenter of liver regeneration
PT proteins - useful for, e.g. treatment of liver disease and cystic
PT fibrosis
PS Claim 48; Fig 15A-J; 220pp; English.
CC The present sequence represents human ATP binding cassette transporter
CC (ABC). The ABC gene is located in the PKD1 locus, between the LCN1 and
CC D16s291 markers in a centromeric to telomeric orientation. The sequence
CC shows homology with murine ABC1 and ABC2 genes. The ABC proteins are
CC responsible for the transport of a wide variety of substrates across
CC cell membranes. Proteins in this family are linked by strong structural
CC similarities. ABC transporters govern unidirectional transport of
CC molecules into or out of cells and across subcellular membranes. The
CC sequence was isolated using an exon trap. Sequences encoding human netrin
CC (hNET), human ribosomal L3 (RPL3L), and human augmenter of liver
CC regeneration (hALR) were also isolated. The antisense oligonucleotides of
CC the DNA sequence are used to modulate expression of ABC prevent its
CC translation. Antibodies against ABC can be used to block binding of its
CC naturally occurring ligands. Host cells containing vectors with DNA
CC inserts encoding the protein can be used in a method for identifying
CC compounds which bind to ABC. Modulation or alteration of hABC3 substrate
CC specificity may have significant therapeutic implications for cystic
CC fibrosis.
SQ Sequence 1704 AA;

Query Match 6.1%; Score 88.5; DB 1; Length 1704;
Best Local Similarity 22.0%; Pred. No. 2.5;
Matches 56; Conservative 30; Mismatches 84; Indels 85; Gaps 12;

```
Qy      39 RKLTSLLSYDPTVPESPDMKFARKRL-----GNLLTTIKHHPSEIIGVLPEDYTRADEEP 93
           | || | :  | :|| |:   | |:      :||     |  ||  | : |
Db    1343 RTLTELYTRMPVLPEDQDVADERTRILAPSPDSLL-----HTPLIIKELSKVY------- 1391

Qy      94 GRQGRPPGRPRKMPRHESSTSLMESPRKTMTRDSKIMFELRG-------KPFEMIAGRFE 146
                      |    |:   | ::       | | |          |:|: |  |
Db    1391 ---------------EQRVPLLAVDRLSLAVQKGECFGLLGFNGAGKTTTFKMLTG--E 1432

Qy     147 EEYSLGRAWVKGH-MNNEYEPIKAQRTDYAP--NLAVDYLACREIHRMPRPDKSIPELPI 203
           |  : | |:| || ::::   :: ||  | |  :  :|::  ||:  |    : |||
Db    1433 ESLTSGDAFVGGHRISSDVGKVR-QRIGYCPQFDALLDHMTGREMLVMYARLRGIPE--- 1489

Qy     204 VPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRR---TAPHARSIALINKIY 260
                                ||      | |     |     |||  :    : |
Db    1489 ------------------------RH-----IGACVENTLRGLLLEPHANKLV---RTY 1515

Qy     261 QPGESKTVEQALGLI 275
           |  : :   : ||
Db    1516 SGGNKRKLSTGIALI 1530
```

RESULT 8
R97614
ID R97614 standard; Protein; 536 AA.
AC R97614;
DT 21-NOV-1996 (first entry)
DE Rat N-acetylglucosaminyl transferase-III, inhibitor of metastasis.
KW N-acetylglucosaminyl transferase-III; metastasis inhibitor; cancer;
KW neoplasia; invasive.
OS Rattus sp.
PN J08109139-A.
PD 30-APR-1996.
PF 12-OCT-1994; 271802.
PR 12-OCT-1994; JP-271802.
PA (TAKI) TAKARA SHUZO CO LTD.
DR WPI; 96-263788/27.
DR N-PSDB; T29198.
PT Inhibitor of cancer metastasis - contains N-acetyl:glucosaminyl
PT transferase-III isolated from rat kidneys
PS Claim 3; Page 6-8; 12pp; Japanese.
CC R97614 and R97615 are two clones of rat N-acetylglucosaminyl
CC transferase-III (GnT-III). These clones are useful to reinforce
CC the activity of GnT-III present in cancer cells and cells surrounding
CC cancer cells for the inhibition of metastasis, preventing the spread
CC of cancer cells to tissues other than that of their origin. The
CC clones were isolated from a rat kidney homogenate.
SQ Sequence 536 AA;

Query Match 6.0%; Score 87.5; DB 1; Length 536;
Best Local Similarity 22.9%; Pred. No. 0.61;
Matches 68; Conservative 30; Mismatches 90; Indels 109; Gaps 18;

```
Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || || :  :  || |  ||  || |||:    |   |  :||     | | :
Db      53 PVTPQASPEPGDPDLLRT-PLYSHSPLLQPL-SPSKATELHRVDFVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
             |              ||  :  ||  |:|     |||  :  :    :|:    |: |
Db     111 AG-------------GVCFKPGTRMLEKPS-----PGRTEEKTKVAEGSSVRGPARRPM 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
             : :   |   | : ||       | ||:    |                 : |:
Db     152 ----RHVLSAR---ERLGGR-----GTRRKWVECVCLPGW-------------HGPSC 184

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------- 213
           |  :   :   :|  :: :|    || |:        |||
Db     185 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAININHEFDLLDVRFHELGDVVDAFVVC 240

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
                |  :||   :   | |   |||| | :            : || |    ||
Db     241 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 296
```

RESULT 9
W24014
ID W24014 standard; Protein; 536 AA.
AC W24014;
DT 10-FEB-1998 (first entry)
DE Rat N-acetylglucosaminyl transferase III.
KW N-acetylglucosaminyl transferase III; GnT-III; rat; human;
KW virus; replication; inhibitor; hepatitis B; hepatitis C; HIV;
KW viral disease; human immunodeficiency virus.
OS Rattus rattus.
PN WO9718836-A1.
PD 29-MAY-1997.
PF 17-JUL-1996; J01986.
PR 17-NOV-1995; JP-322474.
PA (TAKI) TAKARA SHUZO CO LTD.
PI Ihara Y, Miyoshi E, Taniguchi N;
DR WPI; 97-297877/27.
DR N-PSDB; T77093.
PT Virus replication inhibitor containing N-acetyl:glucosaminyl
PT transferase III or its gene - useful in treatment of hepatitis B, C
PT and HIV viral infections
PS Claim 6; Page 17-21; 12pp; Japanese.
CC The present sequence represents rat N-acetylglucosaminyl transferase III
CC (GnT-III). The present specification describes a virus replication
CC inhibitor which contains the gene encoding GnT-III as an active
CC ingredient. The virus replication inhibitor is used against hepatitis B,
CC C and HIV. It is used for the treatment of viral diseases.
SQ Sequence 536 AA;

```
  Query Match             6.0%;  Score 87.5;  DB 1;  Length 536;
  Best Local Similarity   22.9%;  Pred. No. 0.61;
  Matches   68;  Conservative   30;  Mismatches   90;  Indels  109;  Gaps   18;

Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || || :  :  || |  ||  || |||:    |   |  :||     | | :
Db      53 PVTPQASPEPGDPDLLRT-PLYSHSPLLQPL-SPSKATEELHRVDFVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
           |              ||  :  ||  |:|      |||  :  :    :|:    |: |
Db     111 AG-------------GVCFKPGTRMLEKPS-----PGRTEEKTKVAEGSSVRGPARRPM 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
             : :   |    | : ||        | ||:     |                : |:
Db     152 ----RHVLSAR----ERLGGR-----GTRRKWVECVCLPGW-------------HGPSC 184

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
           |  :   :   :|  :: :|    || |:       |||
Db     185 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAININHEFDLLDVRFHELGDVVDAFVVC 240

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
                |  :||   :   | |   |||||  | :           : || |    ||
Db     241 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 296


RESULT  10
R87628
ID   R87628 standard; Protein; 1257 AA.
AC   R87628;
DT   02-AUG-1996  (first entry)
DE   Alpha-ketoglutaric acid dehydrogenase.
KW   L-glutamic acid; L-lysine; production; Coryneform bacteria;
KW   alpha-ketoglutaric acid dehydrogenase; efficient.
OS   Brevibacterium lactofermentum.
PN   WO9534672-A.
PD   21-DEC-1995.
PF   07-JUN-1995; J01131.
PR   14-JUN-1994; JP-131744.
PA   (AJIN ) AJINOMOTO CO INC.
PI   Abe C,  Asakura Y,  Kawahara Y,  Kimura E,  Kurahashi O;
PI   Nakamatsu T, Tsujimoto N,  Usuda Y;
DR   WPI; 96-049699/05.
DR   N-PSDB; T08998.
PT   Coryneform L-glutamic acid producing bacteria - useful in producing
PT   L-glutamic acid and L-lysine
PS   Claim 4; Page 35-44; 62pp; Japanese.
CC   The present sequence is that of an alpha-ketoglutaric acid dehydrogenase
CC   derived from L-glutamic acid producing Coryneform bacteria. L-glutamic
CC   producing Coryneform bacteriaare useful in producing L-glutamic acid.
CC   Coryneform bacteria contg. the DNA and L-lysine producing ability lead
CC   to the productionof L-lysine.
SQ   Sequence   1257 AA;


  Query Match             6.0%;  Score 87;  DB 1;  Length 1257;
  Best Local Similarity   22.7%;  Pred. No. 2.3;
  Matches   44;  Conservative   26;  Mismatches   74;  Indels   50;  Gaps    8;

Qy      20 ARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLGNLLTTIKHHPSEII 79
           |: : |  |  | |  :|: |    :  |    :|  |          :::   |
Db      87 AQPSAPKESAKPAPKAAPAAKAAPRVETKPAAKTAPKAK---------ESSVPQQPK--- 135

Qy      80 GVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTMTRD--SKIMFELRGKP 137
             |||            |: | |       ::    :| |  |  ||  :::||| |
Db     135 --LPE-----------PGQTPIRGIFKSIAKNMDISLEIPTATSVRDMPARLMFENRAMV 181

Qy     138 FEMI----AGRFEEEYSLGRAWVKG-----HMNNEYEPIKAQRTDYAP-----NLAVDYL 183
             : :     |:    : :| | ||       ||| |: |  : |   |      ||:|
Db     182 NDQLKRTRGGKISFTHIIGYAMVKAVMAHPDMNNSYDVIDGKPTLIVPEHINLGLAID-- 240

Qy     184 ACREIHRMPRPDKS 197
                :|: | |
Db     240 -------LPQKDGS 246


RESULT  11
W41781
ID   W41781 standard; Protein; 1257 AA.
AC   W41781;
DT   11-JUN-1998  (first entry)
DE   B. lactofermentum alpha-ketoglutaric acid dehydrogenase.
KW   Alpha-ketoglutaric acid dehydrogenase; recombinant production;
KW   L-glutamic acid.
OS   Brevibacterium lactofermentum.
PN   WO9748790-A1.
PD   24-DEC-1997.
PF   04-JUN-1997; J01886.
PR   17-JUN-1996; JP-155575.
PA   (AJIN ) AJINOMOTO KK.
PA   (AJIN ) AJINOMOTO CO INC.
PI   Kawahara Y, Kimura E, Kuwabara Y, Nakamatsu T;
DR   WPI; 98-063131/06.
DR   N-PSDB; V13157.
PT   Improved fermentative production of amino acids - by recombination
PT   of gene with modified form introduced on plasmid containing
PT   temperature sensitive replicon
PS   Example 3; Pages 56-63; 77pp; Japanese.
CC   The present sequence, which is Brevibacterium lactofermentum
CC   alpha-ketoglutaric acid dehydrogenase, was used in the preparation
CC   of a novel microorganism possessing a gene in a form acting
CC   disadvantageously on the production of a target substance by
CC   fermentation, modified by the introduction of a plasmid containing
CC   a functional version of the gene together with a temperature
CC   sensitivity replicon. The gene is preferably a gene encoding an
CC   enzyme produced by a species capable of biosynthesising the target
CC   substance, and the microorganism containing the plasmid is cultured
CC   at a temperature allowing the plasmid to replicate, then at a
CC   temperature where plasmid replication does not occur so that the
CC   plasmid is therefore detached from the microbial cell.
CC   The microorganism can be used for the efficient production of amino
CC   acids and other target substances by fermentation with improved
CC   yield, in particular, L-glutamic acid production is enhanced using
CC   alpha-ketoglutaric acid dehydrogenase gene, and L-lysine production
CC   using homoserine dehydrogenase gene.
SQ   Sequence   1257 AA;


  Query Match             6.0%;  Score 87;  DB 1;  Length 1257;
  Best Local Similarity   22.7%;  Pred. No. 2.3;
  Matches   44;  Conservative   26;  Mismatches   74;  Indels   50;  Gaps    8;

Qy      20 ARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLGNLLTTIKHHPSEII 79
           |: : |  |  | |  :|: |    :  |    :|  |          :::   |
Db      87 AQPSAPKESAKPAPKAAPAAKAAPRVETKPAAKTAPKAK---------ESSVPQQPK--- 135

Qy      80 GVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTMTRD--SKIMFELRGKP 137
             |||            |: | |       ::    :| |  |  ||  :::||| |
Db     135 --LPE---------PGQTPIRGIFKSIAKNMDISLEIPTATSVRDMPARLMFENRAMV 181

Qy     138 FEMI----AGRFEEEYSLGRAWVKG-----HMNNEYEPIKAQRTDYAP-----NLAVDYL 183
             : :     |:    : :| | ||       ||| |: |  : |   |      ||:|
Db     182 NDQLKRTRGGKISFTHIIGYAMVKAVMAHPDMNNSYDVIDGKPTLIVPEHINLGLAID-- 240

Qy     184 ACREIHRMPRPDKS 197
                :|: | |
Db     240 -------LPQKDGS 246


RESULT  12
R29571
ID   R29571 standard; Protein; 675 AA.
AC   R29571;
DT   20-APR-1993  (first entry)
DE   Human heregulin-beta-1 deduced from clone lambda her11.1dbl.
KW   p185-HER2 ligand; her2 proto-oncogene; breast cancer; EGF receptor;
KW   epidermal growth factor; tyrosine kinase-like glycoprotein;
```

```
KW   HRG-beta1.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   domain          278. .300
FT                   /label= transmembrane
FT                   /note= "probable"
FT   region          212. .251
FT                   /label= EGF_motif
FT                   /note= "contains 6 cysteines"
FT   modified_site   28. .29
FT                   /note= "serine-glycine dipeptide potential
FT                   glycosaminoglycan addition site -
FT                   not part of mature HRG-alpha sequence"
FT   modified_site   50. .51
FT                   /note= "serine-glycine dipeptide potential
FT                   glycosaminoglycan addition site"
FT   modified_site   137. .138
FT                   /note= "serine-glycine dipeptide potential
FT                   glycosaminoglycan addition site"
FT   modified_site   150. .152
FT                   /note= "N-linked glycosylation site"
FT   modified_site   156. .158
FT                   /note= "N-linked glycosylation site"
FT   modified_site   196. .198
FT                   /note= "N-linked glycosylation site"
FT   modified_site   428. .430
FT                   /note= "N-linked glycosylation site"
FT   modified_site   600. .602
FT                   /note= "N-linked glycosylation site"
FT   modified_site   195. .207
FT                   /note= "serine-threonine potential
FT                   O-glycosylation sites"
FT   region          1. .30
FT                   /note= "deduced from ORF but initiating MET is
FT                   at position 31 and the processed
FT                   N-terminal residue is Ser32"
PN   WO9220798-A.
PD   26-NOV-1992.
PF   21-MAY-1992; U04295.
PR   24-MAY-1991; US-705256.
PR   25-SEP-1991; US-765212.
PR   08-NOV-1991; US-790801.
PR   06-MAR-1992; US-847743.
PR   11-MAY-1992; US-880917.
PA   (GENE-) GENETECH INC.
PI   Holmes WE, Vandlen RL;
DR   WPI; 92-415776/50.
DR   N-PSDB; Q31542.
PT   New polypeptide(s) heregulin(s) bind the P185HER2 receptor - for
PT   diagnosing and treating tumours and screening for agonists or
PT   antagonists for binding to P185HER2
PS   Example 6; Fig 8; 139pp; English.
CC   Clone lambda her11.1dbl (heregulin-beta-1) was identified in a
CC   lambda gt10 oligo-dT primed cDNA library derived from human breast
CC   carcinoma MDA-MB-231 cells. Labelled synthetic DNA probes
CC   corresponding to the 5' and 3' ends of lambda her16 (see Q31541)
CC   were used in the hybridisation reaction under high stringency
CC   conditions. The HRG-beta-1 clone extends 189 bp longer than lambda
CC   her16 in the 3' direction and supplies a stop codon after Val 675.
CC   At nucleotide position 205 of lambda her11.1dbl there is an A
CC   substituted for G which results in substitution of Gln in place of
CC   Arg at position 68 in HRG-beta-1. The other differences occur in
CC   the EGF motif of the two HRG proteins.
CC   See also Q31543-Q31548.
SQ   Sequence   675 AA;

  Query Match             5.9%;  Score 86.5;  DB 1;  Length 675;
  Best Local Similarity   20.7%;  Pred. No. 1.1;
  Matches   67;  Conservative   40;  Mismatches  126;  Indels   91;  Gaps   16;

Qy      18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
           |: | |   |: ||     |   :         :::| :  |    :  |:       |

Db     318 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 373

Qy      70 TIKHHPSEII----------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
           :  || : :                :| | ::    :  |  |   | | ||
Db     374 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 433

Qy     105 ----KMPRH--ESSTSLMESPRK--------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
                 ||  |: |  | :||          |  | | : |      :  | ||
Db     434 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 493

Qy     147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
             |:    |   |  |  |        :: :: |: |     :      |  |  | :
Db     494 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 546

Qy     199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
             ||  | || |   |  |||    ||    : |||   : : :|| |:    : |:
Db     546 -SLPASPLRIVE----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 594

Qy     259 I---------YQPGESKTVEQALG 273
           :          ||:| :: :|
Db     595 LEVDSNTSSQSSNSESETEDERVG 618


RESULT  13
W74487
ID   W74487 standard; protein; 675 AA.
AC   W74487;
DT   20-NOV-1998  (first entry)
DE   Amino acid sequence of the heregulin-beta-1 variant 2.
KW   Heregulin variant; ErbB; epidermal growth factor receptor gene;
KW   ErbB receptor; glial cell; Schwann cell; muscle cell; cancer; antibody;
KW   ion channel; acetylcholine receptor; synaptic junction; nerve damage;
KW   oestrogen receptor; aetiology; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; immunogen.
OS   Homo sapiens.
PN   WO9835036-A1.
PD   13-AUG-1998.
PF   10-FEB-1998; U01579.
PR   10-FEB-1997; US-799054.
PA   (GETH ) GENENTECH INC.
PI   Ballinger MD, Fairbrother WJ, Jones JT, Sliwkowski MX,
PI   Wells JA;
DR   WPI; 98-447233/38.
PT   New variants of heregulin with increased affinity for ErbB receptors
PT   - used to treat cancer, nervous system diseases, neuropathy etc. and
PT   for diagnosis
PS   Claim 10; Page -; 160pp; English.
CC   This is the amino acid sequence of a heregulin variant with increased
CC   affinity for ErbB (epidermal growth factor receptor gene) receptors,
CC   used in the method of the invention.  Host cells are used to produce
CC   the heregulin variants, activate cellular ErbB receptors, in vivo or
CC   in cell cultures (e.g. being used to produce cell-specific products or
CC   intended for transplantation), so increase survival, proliferation and
CC   differentiation of cells, especially glial, Schwann and muscle cells.
CC   Heregulin variants are particularly used to treat cancers and other
CC   disorders of the nervous system, muscles and epithelia, but also to
CC   induce formation of ion channels, acetylcholine receptors and synaptic
CC   junctions, downregulate the oestrogen receptor and cause cell
CC   internalisation.  It is also used to treat or prevent nerve nerve damage
CC   or degeneration of any aetiology, in cases such as e.g. Alzheimer's,
CC   Parkinson's, and Huntington's diseases.  They are also used to detect
CC   presence of ErbB receptors by binding assay, particularly
CC   overexpression as diagnostic of cancer, or as immunogens for raising
CC   antibodies.
CC   This sequence was created using the human heregulin sequence from patent
CC   number WO92/20798, and information given in the specification.
SQ   Sequence   675 AA;

  Query Match             5.9%;  Score 86.5;  DB 1;  Length 675;
  Best Local Similarity   20.7%;  Pred. No. 1.1;
  Matches   67;  Conservative   40;  Mismatches  126;  Indels   91;  Gaps   16;
```

```
Qy      18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
           |: | |   |: ||     |   :          :::| :  |    :  |:        |
Db     318 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 373

Qy      70 TIKHHPSEII-------------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
           :  || : :               :| | ::   :  |  |   | | ||
Db     374 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 433

Qy     105 ----KMPRH--ESSTSLMESPRK--------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
               ||  |:  |  :||          |  | | : |      :  | ||
Db     434 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 493

Qy     147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
           |:    |   |  | |         :: :: |: |     :      |  |  | :
Db     494 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 546

Qy     199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
           ||  | || |     |  |||    ||    : |||   : : :|| |:     : |:
Db     546 -SLPASPLRIVE-----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 594

Qy     259 I---------YQPGESKTVEQALG 273
           :             ||:| :: :|
Db     595 LEVDSNTSSQSSNSESETEDERVG 618
```

RESULT 14
W74488
ID W74488 standard; protein; 675 AA.
AC W74488;
DT 20-NOV-1998 (first entry)
DE Amino acid sequence of the heregulin-beta-1 variant 3.
KW Heregulin variant; ErbB; epidermal growth factor receptor gene;
KW ErbB receptor; glial cell; Schwann cell; muscle cell; cancer; antibody;
KW ion channel; acetylcholine receptor; synaptic junction; nerve damage;
KW oestrogen receptor; aetiology; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; immunogen.
OS Homo sapiens.
PN WO9835036-A1.
PD 13-AUG-1998.
PF 10-FEB-1998; U01579.
PR 10-FEB-1997; US-799054.
PA (GETH) GENENTECH INC.
PI Ballinger MD, Fairbrother WJ, Jones JT, Sliwkowski MX,
PI Wells JA;
DR WPI; 98-447233/38.
PT New variants of heregulin with increased affinity for ErbB receptors
PT - used to treat cancer, nervous system diseases, neuropathy etc. and
PT for diagnosis
PS Claim 10; Page -; 160pp; English.
CC This is the amino acid sequence of a heregulin variant with increased
CC affinity for ErbB (epidermal growth factor receptor gene) receptors,
CC used in the method of the invention. Host cells are used to produce
CC the heregulin variants, activate cellular ErbB receptors, in vivo or
CC in cell cultures (e.g. being used to produce cell-specific products or
CC intended for transplantation), so increase survival, proliferation and
CC differentiation of cells, especially glial, Schwann and muscle cells.
CC Heregulin variants are particularly used to treat cancers and other
CC disorders of the nervous system, muscles and epithelia, but also to
CC induce formation of ion channels, acetylcholine receptors and synaptic
CC junctions, downregulate the oestrogen receptor and cause cell
CC internalisation. It is also used to treat or prevent nerve nerve damage
CC or degeneration of any aetiology, in cases such as e.g. Alzheimer's,
CC Parkinson's, and Huntington's diseases. They are also used to detect
CC presence of ErbB receptors by binding assay, particularly
CC overexpression as diagnostic of cancer, or as immunogens for raising
CC antibodies.
CC This sequence was created using the human heregulin sequence from patent
CC number WO92/20798, and information given in the specification.
SQ Sequence 675 AA;

Query Match 5.9%; Score 86.5; DB 1; Length 675;
Best Local Similarity 20.7%; Pred. No. 1.1;

Matches 67; Conservative 40; Mismatches 126; Indels 91; Gaps 16;

```
Qy      18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
           |: | |   |: ||     |   :          :::| :  |    :  |:        |
Db     318 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 373

Qy      70 TIKHHPSEII---------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
           :  || : :           :| | ::   :  |  |   | | ||
Db     374 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 433

Qy     105 ----KMPRH--ESSTSLMESPRK--------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
               ||  |:  |  :||          |  | | : |      :  | ||
Db     434 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 493

Qy     147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
           |:    |   |  | |         :: :: |: |     :      |  |  | :
Db     494 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 546

Qy     199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
           ||  | || |     |  |||    ||    : |||   : : :|| |:     : |:
Db     546 -SLPASPLRIVE----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 594

Qy     259 I---------YQPGESKTVEQALG 273
           :             |||:| :: :|
Db     595 LEVDSNTSSQSSNSESETEDERVG 618
```

RESULT 15
W74486
ID W74486 standard; protein; 675 AA.
AC W74486;
DT 20-NOV-1998 (first entry)
DE Amino acid sequence of the heregulin-beta-1 variant 1.
KW Heregulin variant; ErbB; epidermal growth factor receptor gene;
KW ErbB receptor; glial cell; Schwann cell; muscle cell; cancer; antibody;
KW ion channel; acetylcholine receptor; synaptic junction; nerve damage;
KW oestrogen receptor; aetiology; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; immunogen.
OS Homo sapiens.
PN WO9835036-A1.
PD 13-AUG-1998.
PF 10-FEB-1998; U01579.
PR 10-FEB-1997; US-799054.
PA (GETH) GENENTECH INC.
PI Ballinger MD, Fairbrother WJ, Jones JT, Sliwkowski MX,
PI Wells JA;
DR WPI; 98-447233/38.
PT New variants of heregulin with increased affinity for ErbB receptors
PT - used to treat cancer, nervous system diseases, neuropathy etc. and
PT for diagnosis
PS Claim 10; Page -; 160pp; English.
CC This is the amino acid sequence of a heregulin variant with increased
CC affinity for ErbB (epidermal growth factor receptor gene) receptors,
CC used in the method of the invention. Host cells are used to produce
CC the heregulin variants, activate cellular ErbB receptors, in vivo or
CC in cell cultures (e.g. being used to produce cell-specific products or
CC intended for transplantation), so increase survival, proliferation and
CC differentiation of cells, especially glial, Schwann and muscle cells.
CC Heregulin variants are particularly used to treat cancers and other
CC disorders of the nervous system, muscles and epithelia, but also to
CC induce formation of ion channels, acetylcholine receptors and synaptic
CC junctions, downregulate the oestrogen receptor and cause cell
CC internalisation. It is also used to treat or prevent nerve nerve damage
CC or degeneration of any aetiology, in cases such as e.g. Alzheimer's,
CC Parkinson's, and Huntington's diseases. They are also used to detect
CC presence of ErbB receptors by binding assay, particularly
CC overexpression as diagnostic of cancer, or as immunogens for raising
CC antibodies.
CC This sequence was created using the human heregulin sequence from patent
CC number WO92/20798, and information given in the specification.
SQ Sequence 675 AA;

Query Match 5.9%; Score 86.5; DB 1; Length 675;
Best Local Similarity 20.7%; Pred. No. 1.1;
Matches 67; Conservative 40; Mismatches 126; Indels 91; Gaps 16;

```
Qy      18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
           |: | |   |: ||     |   :         :::| :  |    :  |:       |
Db     318 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 373

Qy      70 TIKHHPSEII------------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
           :  || : :               :| | ::    :  |  |   | | ||
Db     374 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 433

Qy     105 ----KMPRH--ESSTSLMESPRK--------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
                 ||  |:  |  :||          |  | | : |      :  | ||
Db     434 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 493

Qy     147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
             |:    |   |  |  |         :: :: |: |     :       |  |  | :
Db     494 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 546

Qy     199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
             ||  | || |     |  |||    ||    : |||   : : :|| |:     : |:
Db     546 -SLPASPLRIVE-----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 594

Qy     259 I---------YQPGESKTVEQALG 273
           :            ||:| :: :|
Db     595 LEVDSNTSSQSSNSESETEDERVG 618
```

Search completed: August 26, 1999, 11:03:44
Job time: 3098 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 1999, 10:12:08 ; Search time 49.56 Seconds
(without alignments)
222.317 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLLPEDGDRNA.....INKIYQPGESKTVEQALGLI 275

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR_60:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	95.5	6.5	531	2	JN0586	beta-1,4-mannosyl-
2	95	6.5	337	2	S76634	hypothetical prote
3	91	6.2	346	2	T01123	hypothetical prote
4	91	6.2	309	2	T02613	hypothetical prote
5	88.5	6.1	267	2	A46122	homeotic protein H
6	88.5	6.1	1704	2	S71363	probable transport
7	88	6.0	839	2	I50590	class I INCENP pro
8	87.5	6.0	536	2	A43415	beta-1,4-mannosyl-
9	87	6.0	844	2	S61112	regulatory protein
10	87	6.0	5262	2	T03454	ALR protein - huma
11	87	6.0	4957	2	T03455	ALR protein - huma
12	86.5	5.9	640	2	A43273	heregulin precurso
13	86.5	5.9	645	2	B43273	heregulin, splice
14	86.5	5.9	1109	2	A40801	phosphoprotein pho
15	86	5.9	610	2	A49082	calcium-dependent
16	85.5	5.8	605	2	S18648	protein kinase wis
17	85.5	5.8	552	2	I38406	neu differentiatio
18	84.5	5.8	1258	2	JC5765	inositol polyphosp
19	84	5.7	2500	1	WMHUE2	HIV-EP2 enhancer-b
20	84	5.7	924	2	A44945	104K microneme-rho
21	84	5.7	877	2	I50591	class II INCENP pr
22	84	5.7	1234	2	T00363	hypothetical prote
23	84	5.7	5762	2	A41819	proline-rich pepti
24	83.5	5.7	268	2	S18814	homeotic protein H
25	83	5.7	288	2	JS0659	homeotic protein H
26	82.5	5.6	390	2	S67439	hypothetical prote
27	82.5	5.6	2441	2	D71623	erythrocyte membra
28	82.5	5.6	637	2	C43273	heregulin precurso
29	82	5.6	1691	1	A44212	genome polyprotein
30	82	5.6	2477	2	S14428	fibronectin precur
31	82	5.6	334	2	E71103	probable aspartate
32	81.5	5.6	505	2	F71885	catalase - Helicob
33	81.5	5.6	365	2	S47657	peroxidase (EC 1.1
34	81.5	5.6	396	2	S31959	acyl-[acyl-carrier
35	81.5	5.6	536	2	JC4362	beta-1,4-mannosyl-
36	81	5.5	1098	2	S38100	hypothetical prote
37	80.5	5.5	2165	1	RRNZA2	genome polyprotein
38	80.5	5.5	1388	2	T00063	hypothetical prote
39	80.5	5.5	412	2	G02099	LYSP100-A - human
40	80	5.5	301	2	A69765	tartrate dehydroge
41	80	5.5	971	2	S44645	F37A4.8 protein -
42	79.5	5.4	425	1	SAVVTL	satellite RNA-enco
43	79.5	5.4	371	2	S28093	hypothetical prote
44	79.5	5.4	385	2	S53052	repB protein - Lac
45	79.5	5.4	431	2	JC4692	C2 toxin component

ALIGNMENTS

RESULT 1
JN0586
beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144
N;Alternate names: beta-D-mannoside beta1,4 N-acetylglucosaminyltransferase III; N-ac
C;Species: Homo sapiens (man)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 03-Mar-1995
C;Accession: JN0586
R;Ihara, Y.; Nishikawa, A.; Tohma, T.; Soejima, H.; Niikawa, N.; Taniguchi, N.
J. Biochem. 113, 692-698, 1993
A;Title: cDNA cloning, expression, and chromosomal localization of human N-acetylgluc
A;Reference number: JN0586
A;Accession: JN0586
A;Molecule type: mRNA
A;Residues: 1-531 <IHA>
A;Experimental source: fetal liver
C;Comment: This enzyme catalyzes the addition of N-acetylglucosamine in beta1-4 linka
ue.
C;Genetics:
A;Map position: 22q13.1
C;Superfamily: UDP-N-acetylglucosamine acyltransferase
C;Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane prot
F;6-21/Domain: transmembrane #status predicted <TMM>
F;169-180/Region: epidermal growth factor-like
F;139,239,257,395/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 6.5%; Score 95.5; DB 2; Length 531;
Best Local Similarity 22.6%; Pred. No. 1.2;
Matches 67; Conservative 27; Mismatches 90; Indels 113; Gaps 17;

```
Qy      6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
          |:     || |  :|  :  || |  ||  ||  ||:    |   |  :||     | | :
Db     53 PVTPQASPEPGGPDLLRT-PLYSHSPLLQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 110

Qy     64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
           |              ||  :  |:  | |     ||||| : |   : :|   ||  :
Db    111 AG-------------GVCFKPGTKMLERP-----PPGRPEEKPEGANGSSARRPPRYLL 151

Qy    124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
           :             |   ||        | ||:     |              : |:
Db    152 SAR-----------ERTGGR-----GARRKWVECVCLPGW-------------HGPSC 180

Qy    179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
           |  :   :   :|  :: :|    || |:        |||
Db    181 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVC 236

Qy    213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
              | :||   :  | |   |||| | :           : || |    ||
Db    237 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 292
```

RESULT 2
S76634
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 16-Dec-1998
C;Accession: S76634
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

```
S.
A;Reference number: S74322; MUID:97061201
A;Accession: S76634
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-337 <KAN>
A;Cross-references: EMBL:D64004; GB:AB001339; NID:g1001701; PID:d1011229; PID:g1001741
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: hypothetical protein 135

  Query Match             6.5%;  Score 95;  DB 2;  Length 337;
  Best Local Similarity   20.0%;  Pred. No. 0.72;
  Matches   63;  Conservative   41;  Mismatches   87;  Indels  124;  Gaps   14;

Qy       4 IDPLAEFLLPEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDP--TVPESPDMKFAR 61
           :||: :|               |:||  :   : : :|  : :     : | : | : ::
Db      59 LDPILQF------------GQLLLSGSTVFFAAETIRLRGITTEQARRSAPYADDRRVSK 106

Qy      62 KRLGNLLTTIKHHPSEIIGVLPED--------YTRADEEPGRQGRPPG--------RPRK 105
            |:        :  :|:  : |||          |  ::  | ||| |         |||
Db     107 TRV--------YTEAELDQLEPEDEPVARNNRRLRGYDDDARSGRPDGYGEAEARPRPRS 158

Qy     106 MPRHESSTSLMESPRKTMTRDSKIMFELRGKPFEMIAGR----------------FEEE 148
             |:   |:   :|| | :| |              |||                 :|::
Db     159 QGRNAPPTN--PNPRPTRSRPS--------------AGRSAPQRPGPAPGYNDNYGYEDD 202

Qy     149 YSLGRAWVKGHMNNEYEPIKAQR-----------TDYAPNLAVDYLACREIHRMPRPDKS 197
           ||    |  |  :  :|  ::|           | ||:           | |||  :
Db     203 YS---GWESGANDVWDDPTPSRRPPTRRPRPEAGNDPAPS-----------RRPRPSNN 247

Qy     198 IP--------------------ELP---IVPSRIDEFDATVDPRYETDLKNEYIRHWKQ 233
           |                     | |   :   ||| | |  |    |  :   |: :|
Db     248 PPNDSFGDRPERNAPRNARPYEDEPPAAYVDYQPIDEADLTPRPTTPEDPAD---RNQEQ 304

Qy     234 VKKGWCAHQRRRTAP 248
           : |    ||   :|
Db     305 SRSGNPRSQRPSRSP 319


RESULT   3
T01123
hypothetical protein (T21L14.22, F24L7.2) - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 19-Feb-1999
C;Accession: T01123; T00784
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, December 1997
A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.
A;Reference number: Z14209
A;Accession: T01123
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-346 <ROU>
A;Cross-references: EMBL:AC003033; NID:g2702261; PID:g2702278
A;Experimental source: cultivar Columbia
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, February 1998
A;Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A;Reference number: Z14204
A;Accession: T00784
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-346 <RO2>
A;Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914690
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 2
A;Introns: 185/3; 196/3; 214/2; 227/3; 253/3; 289/3; 316/2
A;Note: T21L14.22, F24L7.2

  Query Match             6.2%;  Score 91;  DB 2;  Length 346;
  Best Local Similarity   22.5%;  Pred. No. 1.6;
  Matches   61;  Conservative   40;  Mismatches   92;  Indels   78;  Gaps   16;

Qy      18 RNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLGNLLTTIKHHPSE 77
           |  |||   ::   |: |||        | |  |                     :|  :
Db     111 RPVRQNSNSVAD---PVGSPSPG-----GYTPRGP-----------------VYGYHHGQ 145

Qy      78 IIGVLP--EDYTRADEEPGRQ------GRPPGRPRKM-PR------------HESSTSLM 116
           :  |    : ||      :|      |   | |  : ||              | ||::
Db     146 FVSNLDPMNQFMRAAHPQNQQSPQLGSGHMKGVPHFLQPRLVLVSENVYVEATPSPTSIL 205

Qy     117 ESP--RKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPIKAQRTDY 174
           ::   :|  :|| ::  :| :   :  |     ||| |:|::   |  :| ||
Db     206 DNSGHKKARSRDDALVL-VRKRKVRITEG--ASLYSLCRSWLR---NGAHEGIK------ 254

Qy     175 APNLAVDYLACREIHRMPRP---DKSIPELP--IVPSRIDEFD----ATVDPRYETDLKN 225
                 :| : |    :|:|   ||:   ||  :|   | | |     :|    |:||
Db     254 ----RIDMMTC-----LPKPLPVDKTETSLPKDLVEEAICEEDKEDEESVKHLSESDLLK 304

Qy     226 EYIRHWKQVKKGWCAHQRRRTAPHARSIALI 256
           :|   |:|:     : :| | :   :||:
Db     305 RHIDRAKKVRARLREERLKRIARYKARLALL 335


RESULT   4
T02613
hypothetical protein T19L18.8 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C;Accession: T02613
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC T19L18 genomic sequence.
A;Reference number: Z14681
A;Accession: T02613
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-309 <ROU>
A;Cross-references: EMBL:AC004747; NID:g3413696; PID:g3413703
C;Genetics:
A;Map position: 2
A;Note: T19L18.8

  Query Match             6.2%;  Score 91;  DB 2;  Length 309;
  Best Local Similarity   23.1%;  Pred. No. 1.4;
  Matches   42;  Conservative   24;  Mismatches   70;  Indels   46;  Gaps    7;

Qy      13 PEDGDRN------ARQNDPLISGGPLPLES-PSRKLTSLLSYDPT-----VPESPDMKF 59
           |||   |       || | :|  |  |    |  | : :| :||      | |  ::
Db     116 PEDNSNNNFHQPASIEQNQPFLSRSPSVLHRIKSFNLYNYISQEPTNIIEASPPSVTVET 175

Qy      60 ARKRLGNLLTTIKHHPSEIIGVLPEDY----------TRADEEPGRQGRPPGRPRKMPRH 109
            ::::      :|    |    | | |          |::| ||    |||  |:|| :
Db     176 KQEQVQE--QEVKEEQEEEEQSLEEVYSKLNLNHVARTKSDTEPAAGIRPPKLPKKMKKS 233

Qy     110 ESSTS----------LMESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSL 151
           |: |            :: || |   ::    : :   |     ||: :  |
Db     234 ASTKSPFSHFQEDEISVEARRPATVKVPRVTTVEEADEMVDAKADDF---INRFKHQLKL 290

Qy     152 GR 153
           |
Db     291 QR 292


RESULT   5
A46122
homeotic protein Hox-8 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999
C;Accession: A46122
R;Bell, J.R.; Noveen, A.; Liu, Y.H.; Ma, L.; Dobias, S.; Kundu, R.; Luo, W.; Xia, Y.;
Genomics 16, 123-131, 1993
```

A;Title: Genomic structure, chromosomal location, and evolution of the mouse Hox 8 gene.
A;Reference number: A46122; MUID:93252368
A;Accession: A46122
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-267 <BEL>
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIN:131241, NCBIN:131280, NCBIP:131243)
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation
F;143-199/Domain: homeobox homology <HOX>

Query Match 6.1%; Score 88.5; DB 2; Length 267;
Best Local Similarity 24.3%; Pred. No. 1.9;
Matches 59; Conservative 16; Mismatches 99; Indels 69; Gaps 11;

```
Qy      13 PEDGDRNARQNDPLISGGPLPLES------PSRKLTSLLSYDPTVPESPDMKFARKRLGN 66
           |   : :| :    :|  |  :|:      | ::        | ||  ||   |   |
Db      28 PGGAEGSAEERRVKVSSLPFSVEALMSDKKPPKE-------SPAVP--PDCASAGAVLAP 78

Qy      67 LL-----TTIKHHPSEIIGVLPEDYTRADEEPGRQGRP----PGRPRKMPRHESSTSLM- 117
           ||        :| |  :  | | :      |    | |    |||    ||| | |:
Db      79 LLLPGHGVRDRHSPGPL--VKPFETASVKSENSEDGAPWIQEPGRYSPPPRHMSPTTCTL 136

Qy     117 ------ESPRKTMTRDSKIMFE--LRGKPFEMIAGRFEEEYSLG------------RAW 155
                 ||   |    :  |   | | :  || | |   ||             ||
Db     137 RKHKTNRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSSSLNLTETQVKIWFQNRRAK 196

Qy     156 VKGHMNNEYEPIKAQRTDYAPN------------LAVDYLACREIHRMPRPDKSIPELP 202
           |     | | :|       |:               |  | |    ||        | ||
Db     197 AKRLQEAELEKLKMAAKPMLPSGFSLPFPINSPLQAASIYGASYPFHR--------PVLP 248

Qy     203 IVP 205
           | |
Db     249 IPP 251
```

RESULT 6
S71363
probable transport protein ABC-C - human
C;Species: Homo sapiens (man)
C;Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 17-Jul-1998
C;Accession: S71363
R;Klugbauer, N.; Hofmann, F.
FEBS Lett. 391, 61-65, 1996
A;Title: Primary structure of a novel ABC transporter with a chromosomal localization on
A;Reference number: S71363; MUID:96326608
A;Accession: S71363
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1704 <KLU>
A;Cross-references: EMBL:X97187; NID:g1514529; PID:e243436; PID:g1514530
A;Experimental source: cell line medullary thyroid carcinoma
C;Superfamily: ATP-binding cassette homology
C;Keywords: ATP binding; P-loop; phosphoprotein; transmembrane protein
F;255-283/Domain: transmembrane #status predicted <TM1>
F;307-329/Domain: transmembrane #status predicted <TM2>
F;345-364/Domain: transmembrane #status predicted <TM3>
F;373-394/Domain: transmembrane #status predicted <TM4>
F;401-422/Domain: transmembrane #status predicted <TM5>
F;452-475/Domain: transmembrane #status predicted <TM6>
F;549-739/Domain: ATP-binding cassette homology <ABC1>
F;566-573,1416-1423/Region: nucleotide-binding motif A (P-loop)
F;685-690,1535-1540/Region: nucleotide-binding motif B
F;1100-1120/Domain: transmembrane #status predicted <TM7>
F;1145-1169/Domain: transmembrane #status predicted <TM8>
F;1181-1207/Domain: transmembrane #status predicted <TM9>
F;1215-1236/Domain: transmembrane #status predicted <TM10>
F;1245-1264/Domain: transmembrane #status predicted <TM11>
F;1299-1324/Domain: transmembrane #status predicted <TM12>
F;1399-1590/Domain: ATP-binding cassette homology <ABC2>
F;674,866,1524/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #stat
F;1344/Binding site: phosphate (Thr) (covalent) (by cAMP-dependent kinase) #status pr

Query Match 6.1%; Score 88.5; DB 2; Length 1704;
Best Local Similarity 22.0%; Pred. No. 21;
Matches 56; Conservative 30; Mismatches 84; Indels 85; Gaps 12;

```
Qy      39 RKLTSLLSYDPTVPESPDMKFARKRL-----GNLLTTIKHHPSEIIGVLPEDYTRADEEP 93
           | || | :  | :|  |:   | |:      :||     |   ||  | : |
Db    1343 RTLTELYTRMPVLPEDQDVADERTRILAPSPDSLL-----HTPLIIKELSKVY------- 1391

Qy      94 GRQGRPPGRPRKMPRHESSTSLMESPRKTMTRDSKIMFELRG-------KPFEMIAGRFE 146
                  |    |:   | ::       | | |          |:|: |  |
Db    1391 ---------------EQRVPLLAVDRLSLAVQKGECFGLLGFNGAGKTTTFKMLTG--E 1432

Qy     147 EEYSLGRAWVKGH-MNNEYEPIKAQRTDYAP--NLAVDYLACREIHRMPRPDKSIPELPI 203
           |  : | |:| || :::   :: ||  | |  :  :|::  ||:  |    : |||
Db    1433 ESLTSGDAFVGGHRISSDVGKVR-QRIGYCPQFDALLDHMTGREMLVMYARLRGIPE--- 1489

Qy     204 VPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRR---TAPHARSIALINKIY 260
                                 ||      | |     |     ||| :    : |
Db    1489 -------------------------RH-----IGACVENILRGLLLEPHANKLV---RTY 1515

Qy     261 QPGESKTVEQALGLI 275
           |  : :   : ||
Db    1516 SGGNKRKLSTGIALI 1530
```

RESULT 7
I50590
class I INCENP protein - chicken
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
C;Accession: I50590
R;Mackay, A.M.; Eckley, D.M.; Chue, C.; Earnshaw, W.C.
J. Cell Biol. 123, 373-385, 1993
A;Title: Molecular analysis of the INCENPs (inner centromere proteins): separate doma

A;Reference number: A48482; MUID:94012983
A;Accession: I50590
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-839 <MAC>
A;Cross-references: EMBL:Z25419; NID:g414108; PID:g414109

Query Match 6.0%; Score 88; DB 2; Length 839;
Best Local Similarity 19.0%; Pred. No. 9.3;
Matches 51; Conservative 47; Mismatches 119; Indels 52; Gaps 8;

```
Qy      12 LPEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDM------KFARKRLG 65
           ||   |  : :      |  |||  |            :  || :|:       | | |:
Db    186 LPPSSDDKSPKESSAAESQPLPAAS-----------ELIVPHTPEAKGAGKNKSAFKKTA 234

Qy      66 NLL-TTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTMT 124
           |:  ||:     |:   :| |:  :   |  : |          : :|   |:|  :
Db    235 NVADTTVVLSEKELGLEEVDDSTQVQKHNERDDKEPS--------QRTTDSPETPTGSRL 286

Qy     125 RDSKIMFELRGKPFEMIAGRFEEEYSLGR---AWVKGHMNNEYEPIKAQRTDYAPNLAVD 181
               :   | |||  :     |:||| |   : ::  :        || :     : :|:
Db    287 SRRSVRRSLMGKPSTIRRTSLAEKYSLARKRESTIRKSIARTVIKRKAPQKLSVSSSSVN 346

Qy     182 YLACREIHRMPRPDKSI-----PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKK 236
               |:   |  ::::     | :|  | ::|     :  | :|  :||
Db    347 GSGSEEV---PEDEETVVNAGPPPVPQTPPKLDFQGLRMSLRSQTVNRNE---------- 394

Qy     237 GWCAHQRRRTAPHARSIALINKIYQPGES 265
               |:: |: :   :   |  :| :|
Db    394 -----QQQETSNNECDLSKSEKTQEPPQS 417
```

RESULT 8
A43415

beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144) I
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A43415
R;Nishikawa, A.; Ihara, Y.; Hatakeyama, M.; Kangawa, K.; Taniguchi, N.
J. Biol. Chem. 267, 18199-18204, 1992
A;Title: Purification, cDNA cloning, and expression of UDP-N-acetylglucosamine: beta-D-m
A;Reference number: A43415; MUID:92388193
A;Accession: A43415
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-536 <NIS>
A;Cross-references: GB:D10852; NID:g220821; PID:d1002100; PID:g220822
A;Note: sequence extracted from NCBI backbone (NCBIN:112852, NCBIP:112855)
C;Superfamily: UDP-N-acetylglucosamine acyltransferase
C;Keywords: glycoprotein; glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 6.0%; Score 87.5; DB 2; Length 536;
Best Local Similarity 22.9%; Pred. No. 5.7;
Matches 68; Conservative 30; Mismatches 90; Indels 109; Gaps 18;

```
Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || || :  :  || |  ||  || |||:    |  |  :||     | | :
Db      53 PVTPQASPEPGDPDLLRT-PLYSHSPLLQPL-SPSKATEELHRVDFVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
             |              ||  :  ||  |:|      |||  :  :    :|:    |: |
Db     111 AG-------------GVCFKPGTRMLEKPS-----PGRTEEKTKVAEGSSVRGPARRPM 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
             : :   |    | : ||         | ||:    |              : |:
Db     152 ----RHVLSAR----ERLGGR-----GTRRKWVECVCLPGW-------------HGPSC 184

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------ 213
           |  :   :   :|  :: :|    || |:        |||
Db     185 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAININHEFDLLDVRFHELGDVVDAFVVC 240

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
                |  :||   :   | |   |||| | :           : || |    ||
Db     241 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 296
```

RESULT 9
S61112
regulatory protein GCR1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein LPF10w; protein YPL075w
C;Species: Saccharomyces cerevisiae
C;Date: 23-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 06-Feb-1998
C;Accession: S61112; A25079; A26773
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; Wa
submitted to the EMBL Data Library, August 1995
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A;Reference number: S59677
A;Accession: S61112
A;Molecule type: DNA
A;Residues: 1-844 <HAL>
A;Cross-references: EMBL:U41849; NID:g1147608; PID:g1147618; MIPS:YPL075w
R;Baker, H.V.
Mol. Cell. Biol. 6, 3774-3784, 1986
A;Title: Glycolytic gene expression in Saccharomyces cerevisiae: nucleotide sequence of
A;Reference number: A25079; MUID:87089718
A;Accession: A25079
A;Molecule type: DNA
A;Residues: 1-457,'R',459-844 <BAK>
A;Cross-references: EMBL:M14145; NID:g171589; PID:g171590
R;Holland, M.J.; Yokoi, T.; Holland, J.P.; Myambo, K.; Innis, M.A.
Mol. Cell. Biol. 7, 813-820, 1987
A;Title: The GCR1 gene encodes a positive transcriptional regulator of the enolase and g
A;Reference number: A26773; MUID:87144253
A;Accession: A26773
A;Molecule type: DNA
A;Residues: 1-844 <HOL>
A;Cross-references: EMBL:M15253; NID:g171587; PID:g171588
C;Genetics:
A;Gene: SGD:GCR1
A;Cross-references: SGD:S0005996; MIPS:YPL075w
A;Map position: 16L
C;Keywords: transmembrane protein
F;130-146/Domain: transmembrane #status predicted <TMM>

Query Match 6.0%; Score 87; DB 2; Length 844;
Best Local Similarity 21.7%; Pred. No. 11;
Matches 60; Conservative 32; Mismatches 94; Indels 90; Gaps 13;

```
Qy       2 SEIDPLAEFLLPEDGDRNARQNDPL-------ISGGPL---PL-ESPSRK---------- 41
           : | |:   | |:  |   ::  ||        :| |    |: |||  |
Db     448 NSIQPMLGTLAPQPQDIVQKRKLPLPGSIASAATGSPFSPSPVGESPYSKRFKLDDKPTP 507

Qy      41 ----LTSLLSYDPTVPESPDMKFARKRLGNLLTT--------IKHHPSEIIGVLPEDYTR 88
               | |||:   :  |  |        | |   |         :| |  ::|      |
Db     508 SQTALDSLLTKSISSPRLP-----LSTLANTAVTESFRSPQQFQHSPDFVVGGSSSSTTE 562

Qy      89 ------ADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTMTRD---------------- 127
                 :: |    :   ||| :| ::|:||: ||| :    |
Db     563 NNSKKVNEDSPSSSSKLAERPR-LPNNDSTTSMPESPTEVAGDDVDREKPPESSKSEPND 621

Qy     127 ----SKIMFELRGK-------------PFEMIAGRFEEEYSLGRAWVKG----HMNNEYE 165
               ||   |  ||                |   |    |  | |            ::::|
Db     622 NSPESKDP-EKNGKNSNPLGTDADKPVPISNIHNSTEAANSSGTVTKTAPSFPQSSSKFE 680

Qy     166 PIKAQRTDYAPNLAVDYLACREIHRMPRPDKSIPEL 201
             |  : |   || |: |       :: | :|:| :|
Db     681 IINKKDTKAGPNEAIKY-------KLSRENKTIWDL 709
```

RESULT 10
T03454
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C;Accession: T03454
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Ya
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homo
A;Reference number: Z14954
A;Accession: T03454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-5262 <PRA>
A;Cross-references: EMBL:AF010403; NID:g2358284; PID:g2358285
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Keywords: alternative splicing

Query Match 6.0%; Score 87; DB 2; Length 5262;
Best Local Similarity 22.1%; Pred. No. 1.2e+02;
Matches 67; Conservative 39; Mismatches 99; Indels 98; Gaps 17;

```
Qy      13 PEDGDRNARQNDPLISG-----------GPLPLESPSRKL---------TSL-LSYDPTV 51
           |   | |:||   |: |           | |||:   ::|         ||:  :
Db    2506 PSSMDVNSRQ---LVGGSQAFYQRAPYPGSLPLQQQQQQLWQQQQATAATSMRFAMSARF 2562

Qy      52 PESPDMKFARKRLGNLLTTIKHH-------------PSEII-----------------G 80
           | :|  :  |: ||: |  |                 |:: |                 |
Db    2563 PSTPGPELGRQALGSPLAGISTRLPGPGEPVPGPAGPAQFIELRHNVQKGLGPGGTPFPG 2622

Qy      81 VLPEDYTR---ADEEPGR-----------QGRPPGRPRKMPRHESSTSLMESPR---KTM 123
             |    |     ||:| |           | || :|   |  | : ||  :|     |:
Db    2623 QGPPQRPRFYPVSEDPHRLAPEGLRGLAVSGLPPQKPSAPPAPELNNSLHPTPHTKGPTL 2682

Qy     124 TRDSKIM------FEL-RGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPIKAQRTDYA- 176
              :::       || |  |  : ||:   |          ::::::  || / |
```

```
Db   2683 PTGLELVNRPPSSTELGRPNPLALEAGKLPCE--------DPELDDDFDAHKALEDDEEL 2734

Qy    176 --PNLAVDYL-ACREIHRMPRPDKSIPELPIVPSRIDEFD--ATVDPRYET----DLKNE 226
Db   2735 AHLGLGVDVAKGDDELGTLENLETNDPHLDDLLNG-DEFDLLAYTDPELDTGDKKDIFNE 2793

Qy    227 YIR 229
Db   2794 HLR 2796
```

RESULT 11
T03455
ALR protein - human
C;Species: Homo sapiens (man)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
C;Accession: T03455
R;Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,
Oncogene 15, 549-560, 1997
A;Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
A;Reference number: Z14954
A;Accession: T03455
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-4957 <PRA>
A;Cross-references: EMBL:AF010404; NID:g2358286; PID:g2358287
C;Genetics:
A;Gene: ALR
A;Map position: 12
C;Keywords: alternative splicing

Query Match 6.0%; Score 87; DB 2; Length 4957;
Best Local Similarity 22.1%; Pred. No. 1.1e+02;
Matches 67; Conservative 39; Mismatches 99; Indels 98; Gaps 17;

```
Qy     13 PEDGDRNARQNDPLISG-----------GPLPLESPSRKL---------TSL-LSYDPTV 51
Db   2201 PSSMDVNSRQ---LVGGSQAFYQRAPYPGSLPLQQQQQQLWQQQQATAATSMRFAMSARF 2257

Qy     52 PESPDMKFARKRLGNLLTTIKHH-------------PSEII-----------------G 80
Db   2258 PSTPGPELGRQALGSPLAGISTRLPGPGEPVPGPAGPAQFIELRHNVQKGLGPGGTPFPG 2317

Qy     81 VLPEDYTR---ADEEPGR-----------QGRPPGRPRKMPRHESSTSLMESPR---KTM 123
Db   2318 QGPPQRPRFYPVSEDPHRLAPEGLRGLAVSGLPPQKPSAPPAPELNNSLHPTPHTKGPTL 2377

Qy    124 TRDSKIM------FEL-RGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPIKAQRTDYA- 176
Db   2378 PTGLELVNRPPSSTELGRPNPLALEAGKLPCE--------DPELDDDFDAHKALEDDEEL 2429

Qy    176 --PNLAVDYL-ACREIHRMPRPDKSIPELPIVPSRIDEFD--ATVDPRYET----DLKNE 226
Db   2430 AHLGLGVDVAKGDDELGTLENLETNDPHLDDLLNG-DEFDLLAYTDPELDTGDKKDIFNE 2488

Qy    227 YIR 229
Db   2489 HLR 2491
```

RESULT 12
A43273
heregulin precursor, splice form alpha - human
N;Alternate names: breast cancer cell differentiation factor p45; Neu differentiation fa
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 17-Mar-1999
C;Accession: A43273; A48498; A38155
R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yansur
Science 256, 1205-1210, 1992
A;Title: Identification of heregulin, a specific activator of p185(erbB2).
A;Reference number: A43273; MUID:92271253
A;Accession: A43273
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-640 <HOL>
A;Experimental source: breast tumor cell line, MDA-MB-231, ATCC HTB 26
A;Note: sequence extracted from NCBI backbone (NCBIP:103250)
R;Culouscou, J.M.; Plowman, G.D.; Carlton, G.W.; Green, J.M.; Shoyab, M.
J. Biol. Chem. 268, 18407-18410, 1993
A;Title: Characterization of a breast cancer cell differentiation factor that specifi
A;Reference number: A48498; MUID:93366731
A;Accession: A48498
A;Molecule type: protein
A;Residues: 20-21,'X',23-24,'XX',27-28 <CUL>
R;Peles, E.; Bacus, S.S.; Koski, R.A.; Lu, H.S.; Wen, D.; Ogden, S.G.; Levy, R.B.; Ya
Cell 69, 205-216, 1992
A;Title: Isolation of the neu/HER-2 stimulatory ligand: a 44 kd glycoprotein that ind
A;Reference number: A38155; MUID:92208945
A;Accession: A38155
A;Molecule type: protein
A;Residues: 'X',15-16,'X',18-20,'RG',23-24,'GP',27,'E',29,'XP',32-36 <PEL>
A;Note: sequence extracted from NCBI backbone (NCBIP:91347)
C;Genetics:
A;Gene: GDB:HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: EGF homology
C;Keywords: alternative splicing; glycoprotein
F;182-221/Domain: EGF homology <EGF>

Query Match 5.9%; Score 86.5; DB 2; Length 640;
Best Local Similarity 20.7%; Pred. No. 8.8;
Matches 67; Conservative 40; Mismatches 126; Indels 91; Gaps 16;

```
Qy     18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
Db    283 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 338

Qy     70 TIKHHPSEII-------------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
Db    339 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 398

Qy    105 ----KMPRH--ESSTSLMESPRK---------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
Db    399 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 458

Qy    147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
Db    459 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 511

Qy    199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
Db    511 -SLPASPLRIVE-----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 559

Qy    259 I---------YQPGESKTVEQALG 273
Db    560 LEVDSNTSSQSSNSESETEDERVG 583
```

RESULT 13
B43273
heregulin, splice form beta 1 - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 14-Aug-1998
C;Accession: B43273
R;Holmes, W.E.; Sliwkowski, M.X.; Akita, R.W.; Henzel, W.J.; Lee, J.; Park, J.W.; Yan
Science 256, 1205-1210, 1992
A;Title: Identification of heregulin, a specific activator of p185(erbB2).
A;Reference number: A43273; MUID:92271253
A;Accession: B43273
A;Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A;Molecule type: mRNA
A;Residues: 1-645 <HOL>
C;Genetics:

```
A;Gene: GDB:HGL
A;Cross-references: GDB:132656; OMIM:142445
A;Map position: 8p22-8p11
C;Superfamily: EGF homology
C;Keywords: alternative splicing
F;182-221/Domain: EGF homology <EGF>


  Query Match             5.9%;  Score 86.5;  DB 2;  Length 645;
  Best Local Similarity   20.7%;  Pred. No. 8.9;
  Matches   67;  Conservative   40;  Mismatches  126;  Indels   91;  Gaps   16;

Qy      18 RNARQNDPLISGGPLPLESPSRKL--------TSLLSYDPTVPESPDMKFARKRLGNLLT 69
           |: | |   |: ||     |   :          :::| :  |    :  |:       |
Db     288 RSERNNMMNIANGPHHPNPPPENVQLVNQYVSKNVISSEHIVEREAETSFSTSH----YT 343

Qy      70 TIKHHPSEII------------GVLPEDYT---RADEEPGRQGRPPGRPR--------- 105
           :  || : :                :| | ::    :  |  |.  | | ||
Db     344 STAHHSTTVTQTPSHSWSNGHTESILSESHSVIVMSSVENSRHSSPTGGPRGRLNGTGGP 403

Qy     105 ----KMPRH--ESSTSLMESPRK--------TMTRDSKIMFEL----RGKPFEMIAGRFE 146
                 ||  |:  |  :||          |  | | : |      :  | ||
Db     404 RECNSFLRHARETPDSYRDSPHSERYVSAMTTPARMSPVDFHTPSSPKSPPSEMSPPVSS 463

Qy     147 EEYSLGRAWVKGHMNNEYEP--------IKAQRTDYAPNLAVDYLACREIHRMPRPDKSI 198
           |:    |   |  |  |         :: :: |: |     :       |  |  | :
Db     464 MTVSMPSMAVSPFMEEE-RPLLLVTPPRLREKKFDHHPQQFSSF------HHNPAHDSN- 516

Qy     199 PELPIVPSRIDEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRRTAPHARSIALINK 258
             ||  | || |     |  |||    ||    : |||   : : :|| |:     : |:
Db     516 -SLPASPLRIVE-----DEEYET--TQEYEPAQEPVKKLANSRRAKRTKPNGH---IANR 564

Qy     259 I---------YQPGESKTVEQALG 273
           :          ||:| :: :|
Db     565 LEVDSNTSSQSSNSESETEDERVG 588


RESULT  14
A40801
phosphoprotein phosphatase (EC 3.1.3.16) 1 glycogen-binding regulatory chain - rabbit
N;Alternate names: protein phosphatase-1(G)
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 14-Nov-1997
C;Accession: A40801; S04004
R;Tang, P.M.; Bondor, J.A.; Swiderek, K.M.; DePaoli-Roach, A.A.
J. Biol. Chem. 266, 15782-15789, 1991
A;Title: Molecular cloning and expression of the regulatory (R-G1) subunit of the glycog
A;Reference number: A40801; MUID:91340717
A;Accession: A40801
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1109 <TAN>
A;Cross-references: GB:M65109; NID:g165692; PID:g165693
R;Dent, P.; Campbell, D.G.; Hubbard, M.J.; Cohen, P.
FEBS Lett. 248, 67-72, 1989
A;Title: Multisite phosphorylation of the glycogen-binding subunit of protein phosphatas
A;Reference number: S04004; MUID:89252053
A;Accession: S04004
A;Status: preliminary
A;Molecule type: protein
A;Residues: 33-68 <DEN>
C;Keywords: phosphoprotein; phosphoric monoester hydrolase


  Query Match             5.9%;  Score 86.5;  DB 2;  Length 1109;
  Best Local Similarity   24.1%;  Pred. No. 18;
  Matches   51;  Conservative   24;  Mismatches   84;  Indels   53;  Gaps    9;

Qy      15 DGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMK----FARKRLGN---L 67
           | :||    :  : |    :|:  | |||| : | |:|    |:     :|  ||    :
Db     536 DQERNVGHLEITVEG----IEASDRDLTSLPTKDTTIPTWAIMEDTFHSSRTPLGREEAV 591

Qy      68 LTTIKHHPSE----IIG------------VLPEDYTRADEEPGRQGRPPGRPRKMPRHE 110
           |||  :|  |     |:|              ||  ||    |:       |    |  :|:
Db     592 LTTPEHDLSSSEGTILGGLTGGVCSPRNGNVLKNDYLFQVEKRKSGWINPEDQNKDTQHQ 651

Qy     111 SSTSLMESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPIKAQ 170
           | :::||  |   | ||           || : :|:       | |       :   |:
Db     652 QSWNVLESQEK--ARGSK----------TNIAEQIKEQVDCEDMWEKRDNTGSLKATPAE 699

Qy     171 RTDYAPNLAVDYLACREIHR---MPRPDKSIP 199
                              |:|       |  |  ||
Db     700 A----------LFTCQEAEHCELSPLADHGIP 721


RESULT  15
A49082
calcium-dependent protein kinase isoform AK1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 20-Mar-1998
C;Accession: A49082
R;Harper, J.F.; Binder, B.M.; Sussman, M.R.
Biochemistry 32, 3282-3290, 1993
A;Title: Calcium and lipid regulation of an Arabidopsis protein kinase expressed in E
A;Reference number: A49082; MUID:93213795
A;Contents: ecotype Columbia
A;Accession: A49082
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-610 <HAR>
A;Cross-references: GB:L14771; NID:g289189; PID:g304105
A;Note: sequence extracted from NCBI backbone (NCBIN:128903, NCBIP:128904)
C;Superfamily: calcium-dependent protein kinase; calmodulin repeat homology; protein
C;Keywords: ATP; EF hand; serine/threonine-specific protein kinase
F;148-408/Domain: protein kinase homology <KIN>
F;156-164/Region: protein kinase ATP-binding motif
F;451-483/Domain: calmodulin repeat homology <EF1>
F;487-519/Domain: calmodulin repeat homology <EF2>
F;523-555/Domain: calmodulin repeat homology <EF3>
F;557-589/Domain: calmodulin repeat homology <EF4>


  Query Match             5.9%;  Score 86;  DB 2;  Length 610;
  Best Local Similarity   26.2%;  Pred. No. 9.1;
  Matches   44;  Conservative   23;  Mismatches   59;  Indels   42;  Gaps    9;

Qy      13 PEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLG------- 66
           | ||| :|      :| | :  |: | :| | || :  |    |  :    : |
Db      24 PRDGDDSAS-----MSNGDIASEAVSGELRSRLSDE--VQNKPPEQVTMPKPGTDVETKD 76

Qy      66 -NLLTTIKHHPSEIIGV--LPE--DYTRADEEPGRQGRPPGRPRKMPRH---ESSTSL-M 116
             : |  |    | | :   ||    |::: :|  :  || :|:| |:|    ||  |
Db      77 REIRTESKPETLEEISLESKPETKQETKSETKPESKPDPPAKPKK-PKHMKRVSSAGLRT 135

Qy     117 ESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEY 164
           ||  :  | :                  |:| |||||   :|     :
Db     136 ESVLQRKTEN----------------FKEFYSLGRKLGQGQFGTTF 165


Search completed: August 26, 1999, 11:04:53
Job time: 3165 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 1999, 10:12:10 ; Search time 34.56 Seconds
(without alignments)
224.936 Million cell updates/sec

Title: US-09-087-136-1
Perfect score: 1462
Sequence: 1 MSEIDPLAEFLLPEDGDRNA.....INKIYQPGESKTVEQALGLI 275

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96.5	6.6	678	1	ABPP_RIPCL	Q27905 riptortus c
2	95.5	6.5	531	1	GNT3_HUMAN	Q09327 h beta-1,4-
3	90	6.2	808	1	PLD_TOBAC	P93400 nicotiana t
4	88.5	6.1	1150	1	DPOL_ADECT	P87553 canine aden
5	87.5	6.0	536	1	GNT3_RAT	Q02527 r beta-1,4-
6	87	6.0	844	1	GCR1_YEAST	P07261 saccharomyc
7	86	5.9	610	1	CDP1_ARATH	Q06850 arabidopsis
8	85.5	5.8	605	1	WIS1_SCHPO	P33886 schizosacch
9	85.5	5.8	343	1	Z183_HUMAN	O15541 homo sapien
10	85	5.8	1833	1	ZEP2_HUMAN	P31629 homo sapien
11	84	5.7	924	1	104K_THEPA	P15711 theileria p
12	84	5.7	877	1	INCE_CHICK	P53352 gallus gall
13	83.5	5.7	267	1	HMX2_MOUSE	Q03358 mus musculu
14	83	5.7	288	1	HMXX_CHICK	P50223 gallus gall
15	82.5	5.6	390	1	YD3A_SCHPO	Q10274 schizosacch
16	82	5.6	2477	1	FINC_RAT	P04937 rattus norv
17	82	5.6	1691	1	POLN_HEVME	Q03495 hepatitis e
18	81.5	5.6	536	1	GNT3_MOUSE	Q10470 m beta-1,4-
19	81.5	5.6	396	1	STAD_LINUS	P32062 linum usita
20	81	5.5	2477	1	FINC_MOUSE	P11276 mus musculu
21	80.5	5.5	2165	1	RRPL_HRSVA	P28887 human respi
22	80	5.5	971	1	SN2L_CAEEL	P41877 caenorhabdi
23	80	5.5	301	1	YCSA_BACSU	P42958 bacillus su
24	79.5	5.4	3562	1	PGCV_CHICK	Q90953 gallus gall
25	79.5	5.4	371	1	RLX1_SALTY	P14492 salmonella
26	79	5.4	235	1	BAR6_CHITE	P08726 chironomus
27	79	5.4	621	1	ELL_HUMAN	P55199 homo sapien
28	79	5.4	836	1	GCS1_HUMAN	Q13724 homo sapien
29	79	5.4	2124	1	Y192_HUMAN	Q93074 homo sapien
30	78.5	5.4	555	1	ETV4_MOUSE	P28322 mus musculu
31	78.5	5.4	1935	1	MYSB_PIG	P79293 sus scrofa
32	78	5.3	497	1	IRF5_MOUSE	P56477 mus musculu
33	78	5.3	2004	1	MOZ_HUMAN	Q92794 homo sapien
34	78	5.3	1799	1	YO25_CAEEL	P34675 caenorhabdi
35	77.5	5.3	1183	1	DRPL_RAT	P54258 rattus norv
36	77.5	5.3	920	1	VGLB_HSVSM	Q04464 herpesvirus
37	77.5	5.3	1060	1	ZFH1_DROME	P28166 drosophila
38	77	5.3	510	1	ERM_HUMAN	P41161 homo sapien
39	77	5.3	1051	1	EVI1_HUMAN	Q03112 homo sapien
40	77	5.3	333	1	GSPK_PSEAE	Q00518 pseudomonas
41	77	5.3	436	1	IF4B_YEAST	P34167 saccharomyc
42	77	5.3	527	1	KITH_HSVSA	P21293 herpesvirus
43	77	5.3	2492	1	POLN_EEVVP	P36328 venezuelan
44	77	5.3	1060	1	UAY_EMENI	P49413 emericella
45	77	5.3	802	1	Y250_HUMAN	Q92540 homo sapien

ALIGNMENTS

```
RESULT   1
ABPP_RIPCL
ID   ABPP_RIPCL     STANDARD;      PRT;   678 AA.
AC   Q27905;
DT   01-NOV-1997 (REL. 35, CREATED)
DT   01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PROBABLE ANTIBACTERIAL PEPTIDE POLYPROTEIN PRECURSOR.
OS   RIPTORTUS CLAVATUS.
OC   EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC   PTERYGOTA; HEMIPTERA; EUHEMIPTERA; HETEROPTERA; PANHETEROPTERA;
OC   PENTATOMOMORPHA; COREOIDEA; ALYDIDAE; RIPTORTUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FAT BODY;
RX   MEDLINE; 96261233.
RA   MIURA K., UENO S., KAMIYA K., KOBAYASHI J., MATSUOKA H., ANDO K.,
RA   CHINZEI Y.;
RT   "Cloning of mRNA sequences for two antibacterial peptides in a
RT   hemipteran insect, Riptortus clavatus.";
RL   ZOOL. SCI. 13:111-117(1996).
CC   -!- FUNCTION: HAS ANTIBACTERIAL ACTIVITY IN VITRO.
CC   -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC   -!- SIMILARITY: TO PYRRHOCORICIN, DROSOCIN AND APIDAECIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D49415; G1486237; -.
KW   POLYPROTEIN; INSECT IMMUNITY; ANTIBIOTIC; REPEAT; GLYCOPROTEIN.
FT   DOMAIN        1    678       14 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT        1     67       1-1.
FT   REPEAT       68    114       1-2.
FT   REPEAT      115    161       1-3.
FT   REPEAT      162    208       1-4.
FT   REPEAT      209    255       1-5.
FT   REPEAT      256    302       1-6.
FT   REPEAT      303    349       1-7.
FT   REPEAT      350    396       1-8.
FT   REPEAT      397    443       1-9.
FT   REPEAT      444    490       1-10.
FT   REPEAT      491    537       1-11.
FT   REPEAT      538    584       1-12.
FT   REPEAT      585    631       1-13.
FT   REPEAT      632    678       1-14.
FT   CARBOHYD     32     32       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD     83     83       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    130    130       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    177    177       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    224    224       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    271    271       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    318    318       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    365    365       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    412    412       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    459    459       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    506    506       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    553    553       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    600    600       N-ACETYLGALACTOSAMINE (POTENTIAL).
FT   CARBOHYD    647    647       N-ACETYLGALACTOSAMINE (POTENTIAL).
SQ   SEQUENCE   678 AA;  76367 MW;  E5C3E174 CRC32;
```

```
  Query Match             6.6%;  Score 96.5;  DB 1;  Length 678;
  Best Local Similarity   23.0%;  Pred. No. 1;
  Matches   66;  Conservative   43;  Mismatches  113;  Indels   65;  Gaps   15;

Qy       2 SEIDPLAEF-LLPEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFA 60
           ||: ||:|  :|||  :| :    |:  || ||  :| | :                :
Db     146 SELSPLSEAEVLPEVRERRS----PVDKGGYLPRPTPPRPV----------------YR 184

Qy      61 RKRLGNLLTTIKHHPSEIIGVLPEDYTRAD--EEPGRQGRP-PGRPRKMPRHESSTSLME 117
            :|: :| : :   |     ||||   |    :: |   || | ||    | ::|
Db     185 SRRVASLESELS--PLSEAEVLPEVRERRSPVDKGGYLPRPTPPRPVYRSRRDASLESEL 242

Qy     118 SPRKTMTRDSKIMFEL--RGKPFEM--IAGRFEEEYSLGRAWVKGHMNNEYEPIKAQRTD 173
           ||      : ::: |:  || | :      |      : |:     : :|  |:
Db     243 SP----LSEEEVLPEVRERGSPVDKGGYLPRPTPPRPVYRSRRDASLESELSPL------ 293

Qy     174 YAPNLAVDYLACREIHR-------MPRPDKSIPELPIVPSRID-EFDATVDPRYETDLKN 225
              ::| |    ||          :|||    |  |:  || |   :: : |  | ::
Db     293 ---SVAEDLPEVRERRSPVDKGGYLPRP---TPPRPVYRSRRDASLESELSPLSEAEVLP 346

Qy     226 EYIRHWKQVKKGWCAHQRRRTAPHARSIALINKIYQPGESKTVEQAL 272
           |      | ||   :  | | |          :|:     ::|  |
Db     347 EVRERRSPVDKG--GYLPRPTPP--------RPVYRSRRDASLESEL 383


RESULT   2
GNT3_HUMAN
ID   GNT3_HUMAN     STANDARD;      PRT;   531 AA.
AC   Q09327;
DT   01-NOV-1995 (REL. 32, CREATED)
DT   01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-
DE   TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN
DE   N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE
DE   III) (GNT-III) (GLCNAC-T III).
GN   MGAT3 OR GGNT3.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93380894.
RA   IHARA Y., NISHIKAWA A., TOMA T., SOEJIMA H., NIIKAWA N., TANIGUCHI N.;
RT   "cDNA cloning, expression, and chromosomal localization of human N-
RT   acetylglucosaminyltransferase III (GnT-III).";
RL   J. BIOCHEM. 113:692-698(1993).
CC   -!- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND
CC       BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES
CC       THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE
CC       BETA-LINKED MANNOSE OF THE TRIMANNOSYL CORE OF N-LINKED SUGAR
CC       CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE
CC       REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-
CC       GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-
CC       GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,6)-BETA-D-MANNOSYL-1,4-N-
CC       ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-
CC       1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-
CC       D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-
CC       MANNOSYL-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13789; G398138; -.
DR   PIR; JN0586; JN0586.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; SIGNAL-ANCHOR;
KW   GLYCOPROTEIN; GOLGI STACK.
FT   DOMAIN         1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM       6     21       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN        22    531       LUMENAL, CATALYTIC (POTENTIAL).
FT   DOMAIN        33     84       PRO-RICH.
FT   CARBOHYD     139    139       POTENTIAL.
FT   CARBOHYD     239    239       POTENTIAL.
FT   CARBOHYD     257    257       POTENTIAL.
FT   CARBOHYD     395    395       POTENTIAL.
SQ   SEQUENCE   531 AA;  61026 MW;  E0B3B3B6 CRC32;


  Query Match             6.5%;  Score 95.5;  DB 1;  Length 531;
  Best Local Similarity   22.6%;  Pred. No. 0.92;
  Matches   67;  Conservative   27;  Mismatches   90;  Indels  113;  Gaps   17;

Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || |  :  || |  ||  ||  ||:    |   |  :||     | | :
Db      53 PVTPQASPEPGGPDLLRT-PLYSHSPLLQPL-PPSKAAEELHRVDLVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
             |      ||  :  |:  | |     |||||  : |   : :|    ||  :
Db     111 AG-------------GVCFKPGTKMLERP-----PPGRPEEKPEGANGSSARRPPRYLL 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
           :              |  ||        | ||:     |               : |:
Db     152 SAR-----------ERTGGR-----GARRKWVECVCLPGW-------------HGPSC 180

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------- 213
             |  :   :   :|  :: :|    || |:        |||
Db     181 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAINVNHEFDLLDVRFHELGDVVDAFVVC 236

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
                 |  :||   :   | |   |||| | :           : || |    ||
Db     237 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 292


RESULT   3
PLD_TOBAC
ID   PLD_TOBAC      STANDARD;      PRT;   808 AA.
AC   P93400; P93399;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   PHOSPHOLIPASE D PRECURSOR (EC 3.1.4.4) (PLD) (CHOLINE PHOSPHATASE)
DE   (PHOSPHATIDYLCHOLINE-HYDROLYZING PHOSPHOLIPASE D).
OS   NICOTIANA TABACUM (COMMON TOBACCO).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
OC   ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; NICOTIANA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. SR1; TISSUE=LEAF;
RA   LEIN W., SAALBACH G.;
RL   SUBMITTED (FEB-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELLULAR PATHWAYS INCLUDING
CC       SIGNAL TRANSDUCTION PATHWAYS.
CC   -!- CATALYTIC ACTIVITY: A PHOSPHATIDYLCHOLINE + H(2)O = CHOLINE + A
CC       PHOSPHATIDATE.
CC   -!- COFACTOR: CALCIUM (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z84822; E331460; -.
```

```
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PFAM; PF00168; C2; 1.
DR   PFAM; PF00614; PLDc; 2.
KW   HYDROLASE; LIPID DEGRADATION; CALCIUM.
FT   PROPEP          1     30       BY SIMILARITY.
FT   CHAIN          31    808       PHOSPHOLIPASE D.
FT   DOMAIN          1    109       C2 DOMAIN.
SQ   SEQUENCE   808 AA;  91965 MW;  9CBBF75F CRC32;


  Query Match             6.2%;  Score 90;  DB 1;  Length 808;
  Best Local Similarity   22.6%;  Pred. No. 4.6;
  Matches   55;  Conservative   25;  Mismatches   83;  Indels   80;  Gaps   12;

Qy      65 GNLLTTIKHHPSEIIGV---LPEDYTRADEEPGRQGR------PPGRPRKMP------RH 109
           |:  : || |  | ||     |  |   | |  | ||       |  ||          |
Db      26 GHFFSKIKEHVEETIGFGKGTPAIYATVDLEKARVGRTRKIKNEPNNPRWYESFHIYCAH 85

Qy     110 ESSTSLMESPRKTMTRDSKIMFELRGKPF----EMIAGRFEE----------------- 148
           :|  :      |:  |: |   | |: :    |:: |  ||
Db      86 MASNVIF-----TVKDDNPIGATLIGRAYVPVEELLEG--EEIDKWVEILDREMNPIAEG 138

Qy     148 ----------EYSLGRAWVKGHMNNEYEPIK----AQRTDYAPNLAVDYLACREIHRMPR 193
                    : |    | :|  :::|  :     ||||    :|  |
Db     139 SKIHVKLQFFDVSRDPNWERGIRSSKYPGVPYTFFAQRTGCRVSLYQDAHV--------- 190

Qy     194 PDKSIPELPI------VPSRI--DEFDATVDPRYETDLKNEYIRHWKQVKKGWCAHQRRR 245
           ||  ||::|:       | |   | ||| :: ::       ||  |    :       ||
Db     190 PDNFIPKIPLSGGKYYEPHRCWEDIFDAIINAKHLI-----YITGWSVYTEITLVRDSRR 244

Qy     246 TAP 248
           |
Db     245 QKP 247


RESULT   4
DPOL_ADECT
ID   DPOL_ADECT     STANDARD;      PRT;  1150 AA.
AC   P87553;
DT   15-DEC-1998 (REL. 37, CREATED)
DT   15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   DNA POLYMERASE (EC 2.7.7.7).
GN   POL.
OS   CANINE ADENOVIRUS TYPE 2 (STRAIN TORONTO A 26-61).
OC   VIRUSES; DSDNA VIRUSES, NO RNA STAGE; ADENOVIRIDAE; MASTADENOVIRUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RA   CAMPBELL J.B., ZHAO Y.;
RL   SUBMITTED (NOV-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC       N PYROPHOSPHATE + DNA(N).
CC   -!- THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER.
CC   -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U77082; G1732271; -.
DR   PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW   TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION;
KW   DNA-BINDING.
SQ   SEQUENCE   1150 AA;  130459 MW;  5E7C31F4 CRC32;


  Query Match             6.1%;  Score 88.5;  DB 1;  Length 1150;
  Best Local Similarity   23.2%;  Pred. No. 9.8;
  Matches   45;  Conservative   25;  Mismatches   47;  Indels   77;  Gaps   11;

Qy     110 ESSTSL-MESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEPI- 168
           | :|:| :||| : :| |    |:|:| |      || | |::      ||    :: |
Db     328 EKATTLGLESPEE-LTPDQLKKFKLKGNP------RFIEVYAV------GHNITGFDEIL 374

Qy     168 -KAQRTDYAPNLAVDYLACREIHRMPRPDK--------SIPELPIVPSRIDEFDATVDPR 218
            ||       :  : ||    |||  :         ::|    ||::
Db     375 LAAQVVSTRAEIPPVFEICRNF--MPRAGRLLFNDITYSLPNPSYVPAK----------S 422

Qy     219 YE---------TDLKNEYIRHWKQ--------------------VKKGWCAHQRRRTAPH 249
           ||         :|||::||:   :                    | || | :|
Db     423 YEHWEQGQVLASDLKSQYIKFMVRDTFSLTHTSLKNAAKAYSLTVSKGCCPYQ------- 476

Qy     250 ARSIALINKIYQPG 263
                :|: |  | .
Db     476 -----AVNEFYMLG 484


RESULT   5
GNT3_RAT
ID   GNT3_RAT       STANDARD;      PRT;   536 AA.
AC   Q02527;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   BETA-1,4-MANNOSYL-GLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYL-
DE   TRANSFERASE (EC 2.4.1.144) (N-GLYCOSYL-OLIGOSACCHARIDE-GLYCOPROTEIN
DE   N-ACETYLGLUCOSAMINYLTRANSFERASE III) (N-ACETYLGLUCOSAMINYLTRANSFERASE
DE   III) (GNT-III) (GLCNAC-T III).
GN   MGAT3 OR GNT3.
OS   RATTUS NORVEGICUS (RAT).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 285-295, 445-451 AND 492-507.
RC   STRAIN=DONRYU; TISSUE=KIDNEY;
RX   MEDLINE; 92388193.
RA   NISHIKAWA A., IHARA Y., HATAKEYAMA M., KANGAWA K., TANIGUCHI N.;
RT   "Purification, cDNA cloning, and expression of
RT   UDP-N-acetylglucosamine: beta-D-mannoside
RT   beta-1,4N-acetylglucosaminyltransferase III from rat kidney.";
RL   J. BIOL. CHEM. 267:18199-18204(1992).
CC   -!- FUNCTION: IT IS INVOLVED IN THE REGULATION OF THE BIOSYNTHESIS AND
CC       BIOLOGICAL FUNCTION OF GLYCOPROTEIN OLIGOSACCHARIDES. CATALYZES
CC       THE ADDITION OF N-ACETYLGLUCOSAMINE IN BETA 1-4 LINKAGE TO THE
CC       BETA-LINKED MANNOSE OF THE TRIMANNOSYL CORE OF N-LINKED SUGAR
CC       CHAINS. IT IS ONE OF THE MOST IMPORTANT ENZYMES INVOLVED IN THE
CC       REGULATION OF THE BIOSYNTHESIS OF GLYCOPROTEIN OLIGOSACCHARIDES.
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + N-ACETYL-BETA-D-
CC       GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-
CC       GLUCOSAMINYL-1,2-ALPHA-D-MANNOSYL-1,6)-BETA-D-MANNOSYL-1,4-N-
CC       ACETYL-BETA-D-GLUCOSAMINYL-R = UDP + N-ACETYL-BETA-D-GLUCOSAMINYL-
CC       1,2-ALPHA-D-MANNOSYL-1,3-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,2-ALPHA-
CC       D-MANNOSYL-1,6)-(N-ACETYL-BETA-D-GLUCOSAMINYL-1,4)-BETA-D-
CC       MANNOSYL-1,4-N-ACETYL-BETA-D-GLUCOSAMINYL-R.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10852; G220822; -.
DR   PIR; A43415; A43415.
KW   TRANSFERASE; GLYCOSYLTRANSFERASE; TRANSMEMBRANE; SIGNAL-ANCHOR;
KW   GLYCOPROTEIN; GOLGI STACK.
FT   DOMAIN        1      5       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      6     21       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN       22    536       LUMENAL, CATALYTIC (POTENTIAL).
```

```
FT   DOMAIN       33     84       PRO-RICH.
FT   CARBOHYD    243    243       POTENTIAL.
FT   CARBOHYD    261    261       POTENTIAL.
FT   CARBOHYD    399    399       POTENTIAL.
SQ   SEQUENCE   536 AA;  61762 MW;  FD2FD23A CRC32;


  Query Match           6.0%;  Score 87.5;  DB 1;  Length 536;
  Best Local Similarity  22.9%;  Pred. No. 4.4;
  Matches   68;  Conservative   30;  Mismatches   90;  Indels  109;  Gaps   18;

Qy       6 PLAEFLLPEDGDRNARQNDPLISGGPL--PLESPSRKLTSLLSYDPTVPESPDMKFARKR 63
           |:     || || :  :  || |  ||  || |||:    |   |  :||     | | :
Db      53 PVTPQASPEPGDPDLLRT-PLYSHSPLLQPL-SPSKATEELHRVDFVLPEDTTEYFVRTK 110

Qy      64 LGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTM 123
             |              ||  :  ||  |:|      |||  :  :    :|:    |: |
Db     111 AG-------------GVCFKPGTRMLEKPS-----PGRTEEKTKVAEGSSVRGPARRPM 151

Qy     124 TRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVK-----GHMNNEYEPIKAQRTDYAPNL 178
             : :   |    | : ||        | ||:     |                 : |:
Db     152 ----RHVLSAR----ERLGGR-----GTRRKWVECVCLPGW-------------HGPSC 184

Qy     179 AVDYLACREIHRMPRPDKSIPELPIVPSRI-------DEFD------------------- 213
             |  :   :   :|  :: :|    || |:        |||
Db     185 GVPTVV--QYSNLPTKERLVPR--EVPRRVINAININHEFDLLDVRFHELGDVVDAFVVC 240

Qy     213 -----ATVDPR---YETDLKN---EYIRHWKQV-----------KKGWCAHQRRRT 246
               |  :||   :   | |   ||||| | :           : || |    ||
Db     241 ESNFTAYGEPRPLKFREMLTNGTFEYIRH-KVLYVFLDHFPPGGRQDGWIADDYLRT 296


RESULT   6
GCR1_YEAST
ID   GCR1_YEAST     STANDARD;      PRT;   844 AA.
AC   P07261;
DT   01-APR-1988 (REL. 07, CREATED)
DT   01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   GLYCOLYTIC GENES TRANSCRIPTIONAL ACTIVATOR GCR1.
GN   GCR1 OR SIT3 OR YPL075W OR LPF10W.
OS   SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC   SACCHAROMYCETACEAE; SACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87144253.
RA   HOLLAND M.J., YOKOI T., HOLLAND J.P., MYAMBO K., INNIS M.A.;
RT   "The GCR1 gene encodes a positive transcriptional regulator of the
RT   enolase and glyceraldehyde-3-phosphate dehydrogenase gene families in
RT   Saccharomyces cerevisiae.";
RL   MOL. CELL. BIOL. 7:813-820(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 87089718.
RA   BAKER H.V.;
RT   "Glycolytic gene expression in Saccharomyces cerevisiae: nucleotide
RT   sequence of GCR1, null mutants, and evidence for expression.";
RL   MOL. CELL. BIOL. 6:3774-3784(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   HALL J., AHMED A., BUSSEY H., FORTIN N., FRIESEN J.D.,
RA   STORMS R.K., VO D.H., WANG Y., WINNETT E.;
RL   SUBMITTED (JAN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- FUNCTION: REQUIRED FOR THE EXPRESSION OF MOST GLYCOLYTIC GENES.
CC       GCR1 AND GCR2 PROBABLY FUNCTION AS A TRANSCRIPTIONAL ACTIVATION
CC       COMPLEX, GCR1 PROVIDING THE SPECIFIC DNA-BINDING FUNCTION AND
CC       GCR2 PROVIDING THE ACTIVATION FUNCTION.
CC   -!- SUBUNIT: FORMS A GCR1/GCR2 COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: SOME, TO GCR2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15253; G171588; -.
DR   EMBL; M14145; G171590; -.
DR   EMBL; U41849; G1147618; -.
DR   PIR; A25079; A25079.
DR   SGD; L0000690; GCR1.
DR   TRANSFAC; T00322; -.
KW   TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN.
FT   CONFLICT    458    458       A -> R (IN REF. 2).
SQ   SEQUENCE   844 AA;  94317 MW;  9DE1D2B5 CRC32;


  Query Match           6.0%;  Score 87;  DB 1;  Length 844;
  Best Local Similarity  21.7%;  Pred. No. 8.7;
  Matches   60;  Conservative   32;  Mismatches   94;  Indels   90;  Gaps   13;

Qy       2 SEIDPLAEFLLPEDGDRNARQNDPL-------ISGGPL---PL-ESPSRK---------- 41
           : | |:   | |:  |   ::  ||         :| |    |: |||  |
Db     448 NSIQPMLGTLAPQPQDIVQKRKLPLPGSIASAATGSPFSPSPVGESPYSKRFKLDDKPTP 507

Qy      41 ----LTSLLSYDPTVPESPDMKFARKRLGNLLTT--------IKHHPSEIIGVLPEDYTR 88
               | |||:   : |  |         | |   |          :| |  ::|      |
Db     508 SQTALDSLLTKSISSPRLP-----LSTLANTAVTESFRSPQQFQHSPDFVVGGSSSSTTE 562

Qy      89 ------ADEEPGRQGRPPGRPRKMPRHESSTSLMESPRKTMTRD---------------- 127
                 :: |    :   ||| :| ::|:||: ||| :    |
Db     563 NNSKKVNEDSPSSSSKLAERPR-LPNNDSTTSMPESPTEVAGDDVDREKPPESSKSEPND 621

Qy     127 ----SKIMFELRGK------------PFEMIAGRFEEEYSLGRAWVKG----HMNNEYE 165
               ||   |  ||                |   |    |   | |             ::::|
Db     622 NSPESKDP-EKNGKNSNPLGTDADKPVPISNIHNSTEAANSSGTVTKTAPSFPQSSSKFE 680

Qy     166 PIKAQRTDYAPNLAVDYLACREIHRMPRPDKSIPEL 201
             |  : |   || |: |       :: | :|:| :|
Db     681 IINKKDTKAGPNEAIKY-------KLSRENKTIWDL 709


RESULT   7
CDP1_ARATH
ID   CDP1_ARATH     STANDARD;      PRT;   610 AA.
AC   Q06850;
DT   01-FEB-1995 (REL. 31, CREATED)
DT   01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   CALCIUM-DEPENDENT PROTEIN KINASE, ISOFORM AK1 (EC 2.7.1.-) (CDPK).
GN   AK1.
OS   ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC   EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
OC   EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC   CAPPARALES; BRASSICACEAE; ARABIDOPSIS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE; 93213795.
RA   HARPER J.F., BINDER B.M., SUSSMAN M.R.;
RT   "Calcium and lipid regulation of an Arabidopsis protein kinase
RT   expressed in Escherichia coli.";
RL   BIOCHEMISTRY 32:3282-3290(1993).
CC   -!- FUNCTION: MAY PLAY A ROLE IN SIGNAL TRANSDUCTION PATHWAYS THAT
CC       INVOLVE CALCIUM AS A SECOND MESSENGER.
CC   -!- ENZYME REGULATION: ACTIVATED BY CALCIUM.
CC   -!- ENZYME REGULATION: AUTOPHOSPHORYLATION MAY PLAY AN IMPORTANT
CC       ROLE IN THE REGULATION OF THE KINASE ACTIVITY.
CC   -!- THERE ARE MULTIPLE CDPK ISOFORMS IN MOUSE-EAR CRESS.
CC   -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC       PROTEIN KINASES. BELONGS TO THE CAMK SUBFAMILY.
CC   -!- SIMILARITY: TO EF-HAND CALCIUM BINDING PROTEINS IN THE C-TERMINAL
```

```
CC       SECTION.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L14771; G304105; -.
DR   PIR; A49082; A49082.
DR   PROSITE; PS00018; EF_HAND; 4.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00036; efhand; 4.
DR   PFAM; PF00069; pkinase; 1.
DR   HSSP; P02588; 1PON.
KW   TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW   CALCIUM-BINDING; PHOSPHORYLATION; MULTIGENE FAMILY.
FT   DOMAIN      150    408       PROTEIN KINASE.
FT   NP_BIND     156    164       ATP (BY SIMILARITY).
FT   BINDING     179    179       ATP (BY SIMILARITY).
FT   ACT_SITE    274    274       BY SIMILARITY.
FT   CA_BIND     464    475       POTENTIAL.
FT   CA_BIND     500    511       POTENTIAL.
FT   CA_BIND     536    547       POTENTIAL.
FT   CA_BIND     570    581       POTENTIAL.
SQ   SEQUENCE   610 AA;  68253 MW;  F894E29F CRC32;

  Query Match             5.9%;  Score 86;  DB 1;  Length 610;
  Best Local Similarity   26.2%;  Pred. No. 6.9;
  Matches   44;  Conservative   23;  Mismatches   59;  Indels   42;  Gaps    9;

Qy      13 PEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLG------- 66
           | ||| :|      :| | :  |: | :| | || :  |   |  :    : |
Db      24 PRDGDDSAS-----MSNGDIASEAVSGELRSRLSDE--VQNKPPEQVTMPKPGTDVETKD 76

Qy      66 -NLLTTIKHHPSEIIGV--LPE--DYTRADEEPGRQGRPPGRPRKMPRH---ESSTSL-M 116
            : |  |    | | :   ||    |::: :|  :  || :|:| |:|    ||  |
Db      77 REIRTESKPETLEEISLESKPETKQETKSETKPESKPDPPAKPKK-PKHMKRVSSAGLRT 135

Qy     117 ESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEY 164
           ||  :  | :                 |:| |||||   :|    :
Db     136 ESVLQRKTEN-----------------FKEFYSLGRKLGQGQFGTTF 165


RESULT   8
WIS1_SCHPO
ID   WIS1_SCHPO     STANDARD;      PRT;   605 AA.
AC   P33886;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   PROTEIN KINASE WIS1 (EC 2.7.1.-).
GN   WIS1 OR SPC2.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE; 92097549.
RA   WARBRICK E., FANTES P.A.;
RT   "The wis1 protein kinase is a dosage-dependent regulator of mitosis
RT   in Schizosaccharomyces pombe.";
RL   EMBO J. 10:4291-4299(1991).
CC   -!- FUNCTION: DOSAGE-DEPENDENT REGULATOR OF MITOSIS WITH SERINE/
CC       THREONINE PROTEIN KINASE ACTIVITY. MAY PLAY A ROLE IN THE
CC       INTEGRATION OF NUTRITIONAL SENSING WITH THE CONTROL OVER ENTRY
CC       INTO MITOSIS. IT MAY INTERACT WITH CDC25, WEE1 AND WIN1.
CC       MAY ACTIVATE STY1.
CC   -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
CC       PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62631; G5142; -.
DR   PIR; S18648; S18648.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PFAM; PF00069; pkinase; 1.
DR   HSSP; P24941; 1HCK.
KW   TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW   CELL CYCLE; CELL DIVISION; MITOSIS; PHOSPHORYLATION.
FT   DOMAIN      320    579       PROTEIN KINASE.
FT   NP_BIND     326    334       ATP (BY SIMILARITY).
FT   BINDING     349    349       ATP (BY SIMILARITY).
FT   ACT_SITE    441    441       BY SIMILARITY.
FT   MOD_RES     469    469       PHOSPHORYLATION (BY SIMILARITY).
FT   MOD_RES     473    473       PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   605 AA;  64762 MW;  DD4C56ED CRC32;


  Query Match             5.8%;  Score 85.5;  DB 1;  Length 605;
  Best Local Similarity   25.2%;  Pred. No. 7.5;
  Matches   59;  Conservative   22;  Mismatches   76;  Indels   77;  Gaps   15;

Qy      15 DGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPES-PDMKFAR---KRLGNLLTT 70
           : : :   ||  :|       ||||:: |    || :| : | ::  |    |  | |
Db      57 NSNSSGSDNDSKVS------SPSREIPS----DPPLPRAVPTVRLGRSTSSRSRNSLNL 105

Qy      71 IKHHPSEIIGVLPEDYTRADEEPGRQ----------GRPPGRPRKMPRHESSTSLMESPR 120
              ||             |:| |           | ||  |   |    || : |  :
Db     106 DMKDPS-----------EKPRRSLPTAAGQNNIGSPPTPPGPFP-GGLSTDIQEKLK 150

Qy     121 KTMTRDSKIMFELRGK---PFEMIAGRFEE-EYSLGRAWVKGHMNNEYEPIKAQRTDYAP 176
              | | |:  |   |  :   | |  : : | ::|   |:|      :|
Db     151 AFHASRSKSMPEVVNKISSPTTPIVGMGQRGSYPLPNSQLAGRLSN--SPVK------SP 202

Qy     177 N-----LAVDYLACR-------------EIHRMPRPDK-----SIPELPIVPS 206
           |     ||    | |              | | | | |     | |  |: ||
Db     203 NMPESGLAKSLAAARNPLLNRPTSFNRQTRIRRAP-PGKLDLSNSNPTSPVSPS 255


RESULT   9
Z183_HUMAN
ID   Z183_HUMAN     STANDARD;      PRT;   343 AA.
AC   O15541;
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   ZINC FINGER PROTEIN 183.
GN   ZNF183.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 97368135.
RA   FRATTINI A., FARANDA S., BAGNASCO L., PATROSSO C., NULLI P.,
RA   ZUCCHI I., VEZZONI P.;
RT   "Identification of a new member (ZNF183) of the Ring finger gene
RT   family in Xq24-25.";
RL   GENE 192:291-298(1997).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RA   BRADSHAW H., WU X., OZERSKY P.;
RL   SUBMITTED (AUG-1997) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC   -!- SIMILARITY: CONTAINS A CX(8)CX(5)CX(3)H-TYPE ZINC FINGER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X98253; E247370; -.
DR   EMBL; AC002477; G2341022; -.
DR   PROSITE; PS00518; ZINC_FINGER_C3HC4; 1.
DR   PFAM; PF00097; zf-C3HC4; 1.
DR   PFAM; PF00642; zf-CCCH; 1.
KW   ZINC-FINGER; METAL-BINDING; DNA-BINDING; NUCLEAR PROTEIN.
FT   DOMAIN       86     94       POLY-GLU.
FT   ZN_FING      202    221      CX(8)CX(5)CX(3)H-TYPE.
FT   ZN_FING      262    299      C3HC4-TYPE.
SQ   SEQUENCE   343 AA;  38787 MW;  C21E8EF7 CRC32;


  Query Match             5.8%;  Score 85.5;  DB 1;  Length 343;
  Best Local Similarity   22.3%;  Pred. No. 3.6;
  Matches   69;  Conservative   41;  Mismatches  109;  Indels   91;  Gaps   18;

Qy       4 IDPLAEFLLPEDGDRNA--RQNDPLISGGPLPLESPS---------RKLTSLLSYDPTVP 52
           :| :  ||  : | : |  |:  |     | | || |         |     ::::| :
Db      11 VDQVCTFLFKKPGRKGAAGRRKRPACD--PEPGESGSSSDEGCTVVRPEKKRVTHNPMIQ 68

Qy      53 ESPDMKFARKRLGNLLT-TIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGRPRKMPR--- 109
           :: |    :   |:| :   : :  | :||:        :  |  :| | |  |
Db      69 KTRDSGKQKAAYGDLSSEEEEENEPESLGVV--------YKSTRSAKPVG-PEDMGATAV 119

Qy     109 HESSTSLMESPRKTMTRDSKIMFELRGKPFEMI---AGRFE-----EEYSLGRAWV---- 157
           :|  |      :    |  ||  |||||  : |      ::     :: |:| |
Db     120 YELDTEKERDAQAIFERSQKIQEELRGKEDDKIYRGINNYQKYMKPKDTSMGNASSGMVR 179

Qy     157 KGHMNNEYEPIKAQ-------RTDYAPNLAVDYL---------ACREIHRMPRPDKS-IP 199
           ||       ||:|        | || |::  ||          :|: :|     |:|
Db     180 KG-------PIRAPEHLRATVRWDYQPDICKDYKETGFCGFGDSCKFLH-----DRSDYK 227

Qy     200 ELPIVPSRIDE--FDATVDPRYETDLKNEYI------------------RHWKQVKKGW 238
           :   :||  :    |  ||    :| |                    ||:
Db     228 HGWQIERELDEGRYGVYEDENYEVGSDDEEIPFKCFICRQSFQNPVVTKCRHYFCES--- 285

Qy     239 CAHQRRRTAP 248
           || |  || |
Db     285 CALQHFRTTP 294


RESULT  10
ZEP2_HUMAN
ID   ZEP2_HUMAN     STANDARD;      PRT;  1833 AA.
AC   P31629;
DT   01-JUL-1993 (REL. 26, CREATED)
DT   01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HUMAN IMMUNODEFICIENCY VIRUS TYPE I ENHANCER-BINDING PROTEIN 2
DE   (HIV-EP2).
GN   HIVEP2.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 91217105.
RA   NOMURA N., ZHAO M.-J., NAGASE T., MAEKAWA T., ISHIZAKI R., TABATA S.,
RA   ISHII S.;
RT   "HIV-EP2, a new member of the gene family encoding the human
RT   immunodeficiency virus type 1 enhancer-binding protein. Comparison
RT   with HIV-EP1/PRDII-BF1/MBP-1.";
RL   J. BIOL. CHEM. 266:8590-8594(1991).
RN   [2]
RP   SEQUENCE OF 1184-1323 FROM N.A.
RX   MEDLINE; 91062349.
RA   RUSTGI A.K., VAN'T VEER L.J., BERNARDS R.;
RT   "Two genes encode factors with NF-kappa B- and H2TF1-like DNA-binding
RT   properties.";
RL   PROC. NATL. ACAD. SCI. U.S.A. 87:8707-8710(1990).
CC   -!- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC       5'-GGGACTTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC       NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC       IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC       OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC       MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC       IN T-CELL ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC   -!- SIMILARITY: STRONG, TO HIVEP1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60119; G182120; -.
DR   EMBL; M61744; G187405; -.
DR   PIR; A39829; WMHUE2.
DR   MIM; 143054; -.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2; 2.
DR   PFAM; PF00096; zf-C2H2; 2.
DR   HSSP; P15822; 1BBO.
DR   TRANSFAC; T00939; -.
KW   TRANSCRIPTION REGULATION; ZINC-FINGER; METAL-BINDING; DNA-BINDING;
KW   NUCLEAR PROTEIN.
FT   DOMAIN       324    330      NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN       337    369      SER-RICH.
FT   DOMAIN      1286   1310      ASP/GLU-RICH (ACIDIC).
FT   ZN_FING     1186   1208      C2H2-TYPE.
FT   ZN_FING     1214   1238      C2H2-TYPE.
SQ   SEQUENCE   1833 AA;  202128 MW;  F0A5D053 CRC32;


  Query Match             5.8%;  Score 85;  DB 1;  Length 1833;
  Best Local Similarity   20.1%;  Pred. No. 35;
  Matches   55;  Conservative   39;  Mismatches  124;  Indels   56;  Gaps    8;

Qy      13 PEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDMKFARKRLGNLLTTIK 72
           | |   ::  : |     |    || | |     |    ||    : :|    |: ::
Db    1423 PRDFSPSSHHSSPGYDSSPCRDNSPKRYLI------PKGDLSPRRHLSPRR---DLSPMR 1473

Qy      73 HHPSEIIGVLPEDYTRADEEPGRQGRP-----PGR---------PRKMPRHESSTSLMES 118
           |       |  : :: |  | |    |     ||:         ||:  |:  :|    |
Db    1474 HLSPRKEAALRREMSQRDVSPRRHLSPRRPVSPGKDITARRDLSPRRERRY-MTTIRAPS 1532

Qy     119 PRKTMTRDSKIMF----------------ELRGKPFEMIAGRFEEEYSLGRAWVKGHM 160
           ||: :  :  :                    |   |:  : |  | |    : :
Db    1533 PRRALYHNPPLSMGQYLQAEPIVLGPPNLRRGLPQVPYFSLYGDQEGAYEHPGSSLFPEG 1592

Qy     161 NNEYE----PIKAQRTDYAPNLAVDYLACREIHRMPRPDKSIPELPIVPSRIDEFDATVD 216
           |:|     |: :|:   ||   |     : :| ||    |:   | :|  :: |:
Db    1593 PNDYVFSHLPLHSQQQVRAPIPMVPVGGIQMVHSMPPALSSLHPSPTLPLPMEGFEEKKG 1652

Qy     217 PRYETDLKNEYIREWKQVKKGWCAHQRRRTAPHA 250
             |:  |: |:   :  |:|          |||
Db    1653 ASGESFSKDPYVLSKQHEKRG----------PHA 1676
```

```
RESULT  11
104K_THEPA
ID   104K_THEPA     STANDARD;      PRT;   924 AA.
AC   P15711;
DT   01-APR-1990 (REL. 14, CREATED)
DT   01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT   01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE   104 KD MICRONEME-RHOPTRY ANTIGEN.
OS   THEILERIA PARVA.
OC   EUKARYOTA; ALVEOLATA; APICOMPLEXA; PIROPLASMIDA; THEILERIIDAE;
OC   THEILERIA.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MUGUGA;
RX   MEDLINE; 90158697.
RA   IAMS K.P., YOUNG J.R., NENE V., DESAI J., WEBSTER P.,
RA   OLE-MOIYOI O.K., MUSOKE A.J.;
RT   "Characterisation of the gene encoding a 104-kilodalton microneme-
RT   rhoptry protein of Theileria parva.";
RL   MOL. BIOCHEM. PARASITOL. 39:47-60(1990).
CC   -!- SUBCELLULAR LOCATION: IN MICRONEME/RHOPTRY COMPLEXES.
CC   -!- DEVELOPMENTAL STAGE: SPOROZOITE ANTIGEN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29954; G161866; -.
DR   PIR; A44945; A44945.
KW   ANTIGEN; SPOROZOITE; REPEAT.
FT   DOMAIN        1     19       HYDROPHOBIC.
FT   DOMAIN      905    924       HYDROPHOBIC.
SQ   SEQUENCE   924 AA;  103625 MW;  4563AAA0 CRC32;


  Query Match             5.7%;  Score 84;  DB 1;  Length 924;
  Best Local Similarity   21.9%;  Pred. No. 18;
  Matches   48;  Conservative   27;  Mismatches   78;  Indels   66;  Gaps   10;

Qy       8 AEFLLPEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTV---PESPDMKFARKRL 64
           :||  | :  |       |  :  |||      :|: |  :| |    :|||    | |
Db     730 SEFFTPPESKRTRFHETP--ADTPLP------DVTAELFKEPDVTAETKSPDEAMKRPR- 781

Qy      65 GNLLTTIKHHPSEIIGVLPEDY----------------TRADEEPGRQGR-PPGRPRKM 106
                  |||     | ||                |  : :|||  :   |:| |:
Db     781 ---------SPSEYEDTSPGDYPSLPMKRHRLERLRLTTTEMETDPGRMAKDASGKPVKL 831

Qy     107 PRHESSTSLMESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHMNNEYEP 166
           | :|   |          : |:|: :  |   :      :||                 : |
Db     832 KRSKSFDDLTTVELAPEPKASRIVVDDEGTEAD------DEE-------------THPP 871

Qy     167 IKAQRTDYAPNLAVDYLACREIHRMPRPDKSIPELPIVP 205
           : |:|:            ::  : ||| |  |: |  |
Db     872 EERQKTEVRRRRPP-----KKPSKSPRPSK--PKKPKKP 903


RESULT  12
INCE_CHICK
ID   INCE_CHICK     STANDARD;      PRT;   877 AA.
AC   P53352;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE   INNER CENTROMERE PROTEIN (INCENP).
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 94012983.
RA   MACKAY A.M., ECKLEY D.M., CHUE C., EARNSHAW W.C.;
RT   "Molecular analysis of the INCENPs (inner centromere proteins):
RT   separate domains are required for association with microtubules
RT   during interphase and with the central spindle during anaphase.";
RL   J. CELL BIOL. 123:373-385(1993).
CC   -!- FUNCTION: MAY ACT AS CYTOSKELETAL PROTEINS THAT ARE POTENTIALLY
CC       CAPABLE OF ALTERING THE MORPHOLOGY OF THE CELLULAR MICROTUBULE
CC       NETWORK DURING INTERPHASE.
CC   -!- SUBUNIT: HOMODIMER OR HETERODIMER (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR. IS RESTRICTED TO THE NUCLEUS IN
CC       INTERPHASE, REMAINS TIGHTLY BOUND TO THE CHROMOSOMES UNTIL EARLY
CC       METAPHASE, AND DURING LATE METAPHASE GETS CONCENTRATED IN LINEAR
CC       ARRAYS THAT TRANSECT THE METAPHASE PLATE BETWEEN THE CHROMOSOMES.
CC       AS ANAPHASE BEGINS TO MOVE TO THE SPINDLE MIDZONE WHERE IT IS
CC       INTIMATELY ASSOCIATED WITH THE BUNDLED MICROTUBULES. LATER IN
CC       ANAPHASE GETS CLOSELY ASSOCIATED WITH THE CELL CORTEX, AND BY
CC       TELOPHASE IS CONCENTRATED AT EACH SIDE OF THE MIDBODY IN THE
CC       INTERCELLULAR BRIDGE, WITH WHICH IT IS DISCARDED AFTER.
CC   -!- ALTERNATIVE PRODUCTS: TWO DIFFERENT FORMS; INCENP I AND INCENP II
CC       (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z25419; G414109; -.
DR   EMBL; Z25420; G414111; -.
KW   CELL DIVISION; MICROTUBULES; COILED COIL; CENTROMERE; MITOSIS;
KW   CELL CYCLE; NUCLEAR PROTEIN; ALTERNATIVE SPLICING.
FT   DOMAIN      503    733       COILED COIL (POTENTIAL).
FT   VARIANT     257    257       A -> T.
FT   VARIANT     471    471       E -> Q.
FT   VARSPLIC    716    753       MISSING (IN INCENP I).
SQ   SEQUENCE   877 AA;  100940 MW;  D442B168 CRC32;


  Query Match             5.7%;  Score 84;  DB 1;  Length 877;
  Best Local Similarity   18.5%;  Pred. No. 16;
  Matches   51;  Conservative   47;  Mismatches  112;  Indels   66;  Gaps    8;

Qy      12 LPEDGDRNARQNDPLISGGPLPLESPSRKLTSLLSYDPTVPESPDM------KFARKRLG 65
           ||   |  : :   |   |||  |           :  || :|:       | | |:
Db     186 LPPSSDDKSPKESSAAESQPLPAAS----------ELIVPHTPEAKGAGKNKSAFKKTA 234

Qy      66 NLLTTIKHHPSEIIGVLPEDYT--------RADEEPGRQGRPPGRPRKMPRHESSTSLME 117
           |:  |      : :|:   | :        | |:|| ::                :|   |
Db     235 NVADTTVVLSEKELGLEEVDDSAQVQKHNERDDKEPSQR--------------TTDSPE 279

Qy     118 SPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGR---AWVKGHMNNEYEPIKAQRTDY 174
           :|  :      :  | |||  :      |:||| |   : ::  :       || :
Db     280 TPTGSRLSRRSVRRSLMGKPSTIRRTSLAEKYSLARKRSTIRKSIARTVIKRKAPQKLS 339

Qy     175 APNLAVDYLACREIHRMPRPDKSI-----PELPIVPSRIDEFDATVDPRYETDLKNEYIR 229
             : :|:     |:   |  ::::     | :|  | ::|     :  | :|  :||
Db     340 VSSSSVNGSGSEEV---PEDEETVVNAGPPPVPQTPPKLDFQGLRMSLRSQTVNRNE--- 394

Qy     230 HWKQVKKGWCAHQRRRTAPHARSIALINKIYQPGES 265
                  |:: |: :   ::   |  :| :|
Db     394 -----------QQQETSNNECDLSKSEKTQEPPQS 417


RESULT  13
HMX2_MOUSE
ID   HMX2_MOUSE     STANDARD;      PRT;   267 AA.
AC   Q03358; Q63856;
DT   01-OCT-1993 (REL. 27, CREATED)
```

```
DT   01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   HOMEOBOX PROTEIN MSX-2 (HOX-8.1).
GN   MSX2 OR MSX-2 OR HOX-8.1.
OS   MUS MUSCULUS (MOUSE).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC   RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE; 92037197.
RA   MONAGHAN A.P., DAVIDSON D.R., SIME C., GRAHAM E., BALDOCK R.,
RA   BHATTACHARYA S.S., HILL R.E.;
RT   "The Msh-like homeobox genes define domains in the developing
RT   vertebrate eye.";
RL   DEVELOPMENT 112:1053-1061(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE; 93252368.
RA   BELL J.R., NOVEEN A., LIU Y.H., MA L., DOBIAS S., KUNDU R.,
RA   LUO W., XIA Y.;
RT   "Genomic structure, chromosomal location, and evolution of the mouse
RT   Hox 8 gene.";
RL   GENOMICS 16:123-131(1993).
RN   [3]
RP   SEQUENCE OF 142-202 FROM N.A.
RX   MEDLINE; 91200674.
RA   HOLLAND P.W.H.;
RT   "Cloning and evolutionary analysis of msh-like homeobox genes from
RT   mouse, zebrafish and ascidian.";
RL   GENE 98:253-257(1991).
RN   [4]
RP   FUNCTION.
RX   MEDLINE; 91319208.
RA   DAVIDSON D.R., CRAWLEY A., HILL R.E., TICKLE C.;
RT   "Position-dependent expression of two related homeobox genes in
RT   developing vertebrate limbs.";
RL   NATURE 352:429-431(1991).
CC   -!- FUNCTION: PROBABLE MORPHOGENETIC ROLE. MAY PLAY A ROLE IN LIMB-
CC       PATTERN FORMATION.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN MESENCHYMAL TISSUE IN THE
CC       DEVELOPING SPINAL CORD AND LIMBS.
CC   -!- SIMILARITY: BELONGS TO THE MSH FAMILY OF HOMEOBOX PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L11739; G193966; -.
DR   EMBL; L11738; G193966; JOINED.
DR   EMBL; S60460; E81926; -.
DR   EMBL; S60698; E81926; JOINED.
DR   EMBL; X59252; G57888; -.
DR   EMBL; M38576; G193892; -.
DR   PIR; PS0410; PS0410.
DR   PIR; S18814; S18814.
DR   MGD; MGI:97169; MSX2.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PFAM; PF00046; homeobox; 1.
DR   HSSP; P22808; 1VND.
DR   TRANSFAC; T02076; -.
KW   HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
KW   MULTIGENE FAMILY.
FT   DNA_BIND    142    201       HOMEOBOX.
FT   CONFLICT     89     89       A -> R (IN E81926).
FT   CONFLICT     11     11       F -> FF (IN REF. 1).
FT   CONFLICT    263    263       M -> I (IN REF. 1).
FT   CONFLICT    266    266       L -> V (IN REF. 1).
SQ   SEQUENCE   267 AA;  28915 MW;  B18CB90A CRC32;


  Query Match             5.7%;  Score 83.5;  DB 1;  Length 267;
  Best Local Similarity   24.3%;  Pred. No. 3.8;
  Matches   59;  Conservative   15;  Mismatches  100;  Indels   69;  Gaps   11;

Qy      13 PEDGDRNARQNDPLISGGPLPLES------PSRKLTSLLSYDPTVPESPDMKFARKRLGN 66
           |    : :| :    :|  |  :|:      | ::        | ||  ||   |  .|
Db      28 PGGAEGSAEERRVKVSSLPFSVEALMSDKKPPKE-------SPAVP--PDCASAGAVLRP 78

Qy      67 LL-----TTIKHHPSEIIGVLPEDYTRADEEPGRQGRP----PGRPRKMPRHESSTSLM- 117
           ||          | |  :  | | :      |    | |    |||    ||| | |:
Db      79 LLLPGHGVRDAHSPGPL--VKPFETASVKSENSEDGAPWIQEPGRYSPPPRHMSPTTCTL 136

Qy     117 ------ESPRKTMTRDSKIMFE--LRGKPFEMIAGRFEEEYSLG------------RAW 155
                 ||   |     :  |   | | :  || | |   ||              ||
Db     137 RKHKTNRKPRTPFTTSQLLALERKFRQKQYLSIAERAEFSSSLNLTETQVKIWFQNRRAK 196

Qy     156 VKGHMNNEYEPIKAQRTDYAPN------------LAVDYLACREIHRMPRPDKSIPELP 202
             |     | | :|       |:             |  | |    ||        | ||
Db     197 AKRLQEAELEKLKMAAKPMLPSGFSLPFPINSPLQAASIYGASYPFHR--------PVLP 248

Qy     203 IVP 205
           | |
Db     249 IPP 251


RESULT  14
HMXX_CHICK
ID   HMXX_CHICK     STANDARD;      PRT;   288 AA.
AC   P50223;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HOMEOBOX PROTEIN GHOX-7 (CHOX-7) (HOX-7).
GN   GHOX-7.
OS   GALLUS GALLUS (CHICKEN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC   NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHITE LEGHORN; TISSUE=LIMB BUD;
RX   MEDLINE; 92118001.
RA   NOHNO T., NOJI S., KOYAMA E., NISHIKAWA K., MYOKAI F., SAITO T.,
RA   TANIGUCHI S.;
RT   "Differential expression of two msh-related homeobox genes Chox-7 and
RT   Chox-8 during chick limb development.";
RL   BIOCHEM. BIOPHYS. RES. COMMUN. 182:121-128(1992).
RN   [2]
RP   SEQUENCE OF 130-288 FROM N.A.
RC   TISSUE=LIMB BUD;
RX   MEDLINE; 92283380.
RA   COELHO C.N.D., SUMOY L., KOSHER R.A., UPHOLT W.B.;
RT   "GHox-7: a chicken homeobox-containing gene expressed in a fashion
RT   consistent with a role in patterning events during embryonic chick
RT   limb development.";
RL   DIFFERENTIATION 49:85-92(1992).
CC   -!- FUNCTION: PROBABLY PLAYS A ROLE IN PATTERNING EVENTS DURING
CC       EMBRYONIC LIMB DEVELOPMENT. MAY ALSO BE INVOLVED IN PROGRAMMED
CC       CELL DEATH.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- DEVELOPMENTAL STAGE: EARLY LIMB DEVELOPMENT. FIRST PRESENT IN
CC       AN ASYMMETRIC ARC EXTENDING FROM THE ANTERIOR BORDER OF THE LIMB
CC       BUD TO THE MESENCHYMAL CELLS DIRECTLY ADJACENT TO THE AER. LATER
CC       ABUNDANTLY EXPRESSED IN THE PROXIMAL ANTERIOR PERIPHERY AND
CC       IN THE MID-PROXIMAL REGION OF THE POSTERIOR PERIPHERY. IN OLDER
CC       WING BUDS, DETECTABLE THROUGHOUT THE NECROTIC MESENCHYME BETWEEN
CC       THE DEVELOPING DIGITS.
CC   -!- SIMILARITY: BELONGS TO THE MSH FAMILY OF HOMEOBOX PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D10372; G464147; -.
DR   EMBL; X65791; G396157; -.
DR   PROSITE; PS00027; HOMEOBOX_1; 1.
DR   PROSITE; PS50071; HOMEOBOX_2; 1.
DR   PFAM; PF00046; homeobox; 1.
DR   HSSP; P22808; 1VND.
KW   HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
KW   POLYMORPHISM.
FT   DOMAIN      142    145       POLY-PRO.
FT   DNA_BIND    161    220       HOMEOBOX.
FT   VARIANT     247    247       G -> S (IN GHOX-7D).
FT   CONFLICT    130    130       E -> G (IN REF. 2).
SQ   SEQUENCE   288 AA;  30517 MW;  1C8F7DA1 CRC32;


  Query Match             5.7%;  Score 83;  DB 1;  Length 288;
  Best Local Similarity   24.8%;  Pred. No. 4.7;
  Matches   57;  Conservative   15;  Mismatches  100;  Indels   58;  Gaps   10;

Qy       2 SEIDPLAEFLLPEDG-------DRNARQNDPLISGGPLP-----LESPSRKLTSLLSYDP 49
           |:  | : |  |  |              : | :|  |||     | :  ||       |
Db      16 SDEPPASAFSKPGGGLPVAAAMGGEEESDKPKVSPSPLPFSVEALMADRRKPPG--GRDG 73

Qy      50 TVPESPDMKFARKRLGNLLTTIK--------HHPS-EIIGVLPEDYTRADEEPGRQGR-- 99
             | :  ||  || | |           | ||   :| ||||     | | :  |
Db      74 PEGSGPPLGSARANLGALTTEAPTSPLPLGGHFPSVGALGKLPEDALLKAESPEKPERSP 133

Qy      99 ---------PPGRPRKMP-----RHESSTSLMESPRKTMTRDSKIMFE--LRGKPFEMIA 142
                  || |     |     :|:::      ||   |    :  |   | | :  ||
Db     134 WMQSPRFSPPPPRRLSPPACTLRKHKTN----RKPRTPFTTAQLLALERKFRQKQYLSIA 189

Qy     143 GRFEEEYSLG------------RAWVKGHMNNEYEPIKAQRTDYAPNLA 179
             | |   ||              ||  |     | | :|       |  |
Db     190 ERAEFSSSLSLTETQVKIWFQNRRAKAKRLQEAELEKLKMAAKPMLPPAA 239


RESULT  15
YD3A_SCHPO
ID   YD3A_SCHPO     STANDARD;      PRT;   390 AA.
AC   Q10274;
DT   01-OCT-1996 (REL. 34, CREATED)
DT   01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 43.9 KD PROTEIN C13G7.10 IN CHROMOSOME I.
GN   SPAC13G7.10.
OS   SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC   EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC   SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC   SCHIZOSACCHAROMYCES.
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   CONNOR R., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL   SUBMITTED (FEB-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z69729; E223699; -.
KW   HYPOTHETICAL PROTEIN.
SQ   SEQUENCE   390 AA;  43922 MW;  FCB5FCE4 CRC32;


  Query Match             5.6%;  Score 82.5;  DB 1;  Length 390;
  Best Local Similarity   19.1%;  Pred. No. 7.6;
  Matches   54;  Conservative   40;  Mismatches   87;  Indels  101;  Gaps   13;

Qy      26 LISGGPLPLESPSRKLTSL----LSYDPTVPESPDMKFARKRLGNLLTTIKHH------- 75
           |::|  |  |:||::  :         ||     | :|:  |   :||   : |
Db      21 LLNGSILAAENPSKREVAQDVPGFERKPTKVRKPRVKWTEKETNDLLRGCQIHGVGNWKK 80

Qy      75 ------------PSEIIG----VLPEDYTRADEEPGRQGRPPGRPRKMPRHESSTSLME 117
                        |::::      :|||||  :    |  :    |||:|:|    :  | :
Db      81 ILLDERFHFTNRSPNDLKDRFRTILPEDYKKF--YPNAKTH-MGRPQKIPH---TVGLSK 134

Qy     118 SPRK-----TMTRDSKIM--FELRGKPFEMIA-------------------GRFEEEYS 150
           | ||     |   | :::  | | |  :  |:                        | | |:
Db     135 STRKERKQFTPEEDERLLEGFFLHGPCWTRISKDANLGLQNRRSTDLRDRFRNAFPERYA 194

Qy     151 LGRAWVK---GHMNNEYE--------------------------PIKAQRTDYAPNLAV 180
             :|   |: :  |:                             : |   : :|
Db     195 AAGFKLKNNPGNRSKYYQNNMVNDATTPNDSSTTEAAAAAVAAVAAVAASNPNASPQQTT 254

Qy     181 DYLACREIHRMPRPDKSIPELPIVPSRIDEFDATVDPRYETD 222
           :  |  |:   |  :        :||:   | :  :| | ||
Db     255 EQPASDELLDWPHHN--------LPSQF--FTSQRNPNYSTD 286


Search completed: August 26, 1999, 11:05:32
Job time: 3202 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 26, 1999, 10:11:00 ; Search time 908.98 Seconds
(without alignments)
2949.840 Million cell updates/sec

Title: US-09-087-136-2
Perfect score: 954
Sequence: 1 atgtcagaaatagatccact.....ttcaaatacatcatcttgac 954

Scoring table: IDENTITY_NUC

Searched: 646344 seqs, 1405317327 residues

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	439.2	46.0	32981	36	CELZK418	U00047 Caenorhabdi
	2	46.8	4.9	260699	35	AC006893	AC006893 Caenorhab
c	3	45.2	4.7	228418	35	AC005505	AC005505 Plasmodiu
c	4	43.4	4.5	137791	35	AC006981	AC006981 Homo sapi
c	5	42.2	4.4	85779	7	SCE011856	AJ011856 Saccharom
c	6	41.4	4.3	40460	11	AC005492	AC005492 Homo sapi
c	7	41.2	4.3	110000	34	CEY111B2_1	Continuation (2 of
	8	41.2	4.3	110000	34	CEY37D8_2	Continuation (3 of
	9	41.2	4.3	132742	36	CEY37D8A	AL032626 Caenorhab
c	10	40.8	4.3	35964	36	CEB0365	Z81028 Caenorhabdi
	11	40.2	4.2	4393	2	AF031959	AF031959 Anabaena
	12	40	4.2	176716	35	AC004157	AC004157 Plasmodiu
c	13	40	4.2	176716	35	AC004157	AC004157 Plasmodiu
	14	39.6	4.2	149554	10	HSAC000111	AC000111 Human BAC
c	15	39.6	4.2	202291	34	AC004153	AC004153 Plasmodiu
	16	39.6	4.2	296687	35	AC006713	AC006713 Caenorhab
c	17	39.6	4.2	12029	37	AE001411	AE001411 Plasmodiu
c	18	39.4	4.1	27694	7	HASMT	D31785 Hansenula w
c	19	39.4	4.1	66704	11	AC000015	AC000015 Homo sapi
	20	39.4	4.1	312766	34	PFMAL4P3	AL035476 Plasmodiu
	21	39.4	4.1	118688	35	AC007178	AC007178 Arabidops
c	22	39.4	4.1	6883	36	PFSC04010	AL010216 Plasmodiu
c	23	39.2	4.1	121524	8	AF041468	AF041468 Guillardi
	24	39.2	4.1	89990	11	AC005303	AC005303 Homo sapi
c	25	38.8	4.1	185265	35	AC005412	AC005412 Homo sapi
	26	38.8	4.1	113899	36	PFMAL3P4	AL008970 Plasmodiu
	27	38.4	4.0	200000	34	AC006445	AC006445 *** SEQUE
c	28	38.4	4.0	187647	34	CEY16B4	AL021148 Caenorhab
	29	38.4	4.0	110000	34	CEY39E4_0	Z94158 Caenorhabdi
c	30	38.4	4.0	175374	34	PFMAL1P4	AL031747 Plasmodiu
c	31	38.4	4.0	42734	36	CEDY3	Z96047 Caenorhabdi
	32	38.4	4.0	25305	36	CEF33C8	Z69790 Caenorhabdi
	33	38.4	4.0	33511	36	CEZK1010	Z82083 Caenorhabdi
c	34	38.2	4.0	52120	9	HS217C2	Z82243 Homo sapien
	35	38.2	4.0	35720	34	AC006833	AC006833 Caenorhab
	36	38.2	4.0	202734	34	PFMAL13P1	AL049180 Plasmodiu
	37	38.2	4.0	31430	36	CELC01C4	U41025 Caenorhabdi
	38	38.2	4.0	38655	36	CELC02F12	U41545 Caenorhabdi
	39	38	4.0	82965	34	AC005121	AC005121 Drosophil
c	40	38	4.0	33524	34	AC005122	AC005122 Drosophil
c	41	38	4.0	164399	36	PFMAL3P6	Z98551 Plasmodium
c	42	38	4.0	9304	36	PFSC03105	AL010212 Plasmodiu
c	43	37.8	4.0	7829	1	RPPLAS	X71612 Buchnera ap
	44	37.8	4.0	1865	7	MISCAAP1	X00960 Yeast mitoc
	45	37.8	4.0	758	7	SCAAP1	X00895 Saccharomyc

ALIGNMENTS

RESULT 1
CELZK418/c
LOCUS CELZK418 32981 bp DNA INV 10-MAY-1994
DEFINITION Caenorhabditis elegans cosmid ZK418.
ACCESSION U00047
NID g470368
VERSION U00047.1 GI:470368
KEYWORDS .
SOURCE Caenorhabditis elegans strain=Bristol N2.
ORGANISM Caenorhabditis elegans
Eukaryotae; mitochondrial eukaryotes; Metazoa; Nematoda;
Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea;
Rhabditidae; Caenorhabditis.
REFERENCE 1 (bases 1 to 32981)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A.,
Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans [see comments]

```
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
REFERENCE   2  (bases 1 to 32981)
  AUTHORS   Fulton,L.
  TITLE     The sequence of C. elegans cosmid ZK418
  JOURNAL   Unpublished (1994)
REFERENCE   3  (bases 1 to 32981)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-1994)
COMMENT     Submitted by:
                  Genome Sequencing Center
                  Department of Genetics, Washington University,
                  St. Louis, MO  63110, USA, and
                  Sanger Centre, Hinxton Hall
                  Cambridge CB10 IRQ, England
                  e-mail: rw@nematode.wustl.edu and jes@sanger.ac.uk


          NEIGHBORING COSMID INFORMATION:


          5' overlapping cosmid is R01H2; 3' overlapping cosmid is B0280;
          Actual ZK418 begins at base 197 of this sequence and ends in B0280.


          NOTES:

          Coding sequences below are predicted from computer analysis, using
          the program Genefinder (P. Green and L. Hillier, ms in
          preparation).
FEATURES             Location/Qualifiers
     source          1. .32981
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /map="III"
                     /strain="Bristol N2"
     gene            92. .1634
                     /gene="ZK418.3"
     CDS             join(92. .178,1121. .1234,1283. .1405,1473. .1634)
                     /gene="ZK418.3"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50687.1"
                     /db_xref="PID:g470371"
                     /db_xref="GI:470371"
                     /translation="MADFSCINVCIPCNSCSLRQLLNGCCPRLHSLLSLTYQVIAFSA
                     CLVHLGSEGVRLGLGFYGNLAENMSALFGFLITTIIIQIPLTVFLAVNGSFLNLPLEY
                     VFYTMITVFAIEALMKTAEREVAKYINKIQQEKVADELEKTSGKSPKSGRKTLKKQLS
                     R"
     gene            complement(2055. .3301)
                     /gene="ZK418.4"
     CDS             complement(join(2055. .2228,2276. .2555,2993. .3127,
                     3177. .3301))
                     /gene="ZK418.4"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50688.1"
                     /db_xref="PID:g470372"
                     /db_xref="GI:470372"
                     /translation="MSEIDPLAEFLLPEDGDRNARQNDPLISGGPLPLESPSRKLTSL
                     LSYDPTVPESPDMKFARKRLGNLLTTIKHHPSEIIGVLPEDYTRADEEPGRQGRPPGR
                     PRKMPRHESSTSLMESPRKTMTRDSKIMFELRGKPFEMIAGRFEEEYSLGRAWVKGHM
                     NNEYEPIKAQRTDYAPNLAVDYLACREIHRMPRPDKSIPELPIVPSRIDEFDATVDPR
                     YETDLKNEYIRHWKQVKKG"
     gene            complement(3582. .4582)
                     /gene="ZK418.5"
     CDS             complement(join(3582. .3717,3795. .3904,3955. .4152,
                     4330. .4463,4510. .4582))
                     /gene="ZK418.5"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50689.1"
                     /db_xref="PID:g470373"
                     /db_xref="GI:470373"
                     /translation="MSFFHFINCFALAFAPYFIVYKYSGINEYSSIWKCATASGGYLL
                     TQLAKLLIIATFFPALDSEGFSIVPEFLKSSADIIDVIGLHLLMTNFLAGKGEVRFVV
                     GGLGWGFAHSVAHRLVLLWVGARGTAFTWRWVQTSLDSSADLLVIVSLACLTWMITRT
                     PNKFLVSPILAITVQHTFSLYGWSLLAFRFAYSIATAILTVVVYSANRTASTRKNE"
     gene            complement(5033. .9583)
                     /gene="ZK418.6"
     CDS             complement(join(5033. .5578,5882. .6128,6423. .6777,
                     6943. .7091,7366. .7533,7582. .7789,8102. .8318,8371. .8501,
                     8555. .8718,8764. .8897,9314. .9400,9512. .9583))
                     /gene="ZK418.6"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50690.1"
                     /db_xref="PID:g470374"
                     /db_xref="GI:470374"
                     /translation="MFGEQCSYKKGVGGMIDDIPFIFQNCSQYFVDGDEIYATCDIDK
                     ETDPCYLLKELHRSDTFSTYFLAIIPMVLSIITFIINLGYLIAQIKCFRYETMGFKKR
                     EAFLMSRSSSNIFAIVLFYVILIVWKVNGFNYVSAMIFIIIGGSTFMSITGIYIVLTV
                     ILYMAVAHHISYMTTITLTHCWLIIGLIWIISTVTSVFVGMWGATLFYPDSAPFTCSF
                     ESCQQPLAIVIVIMLSICYGTVIGLYIAMMARIHQLVRKSSMVQARCNSNSMIAMRRL
                     SMNMITFAIGTIPILIVCIVALVNLKELSSLGEGCKSPCKTFRLSYLLIDVEMMASIS
                     AIVWLIAMIADPVINILSDKKLRGVFIKQFSKFPSIFRKSKERTTSITEVSEDHSEVE
                     KFKPQKRRKNERPPRIRRWGKDRLLSHGREKGTLGHQFFFSRPRDITITSRCQRPPER
                     PSTSCKLEAGENLRRRPAPPKQNHVASPAEDQLDTSVWKSKHIRTDVIKAVELGKSPK
                     RVPGLVGKCDCACCGRADDTQTVCEQGGAAKDEYLEEKKRRLQEAKKKRRHHKSTFTL
                     LFYGDIDRKSSEKLKYISSQSNKKRRHRGATRTRSISSPPLPPPAPPNSFMSPDNQTL
                     WCKLYEQTCPPPAKAIDLKMLRNLPGQIGMSELEPEDLEVTPDPPEFIPPRSPGLQIM
                     KPPPVSPPSNIPELIHHQPALVTQPESRPDPNLIILPAPTDVHPPKSPVHHVIKQESM
                     KDIPPGSPASPGAWSGSNSSLIFMPSSTQAPEHYAIPSVSNVLMFRKMEHIFYSSESN
                     WWLLVALVMLTIFAYRSSMEQHTCYYD"
     gene            complement(9825. .11536)
                     /gene="ZK418.7"
     CDS             complement(join(9825. .9969,10015. .10439,10617. .10747,
                     10833. .10996,11042. .11175,11228. .11317,11426. .11536))
                     /gene="ZK418.7"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50691.1"
                     /db_xref="PID:g470375"
                     /db_xref="GI:470375"
                     /translation="MVEVKIYLFFFFRFHFAVHFLFGDGRMSGQSVPLIFLNCSSIFI
                     FGKDEKYAECPIENGTDPCYLLKEVHRAKEFRTYFLAILPIIFSVIACLLNIGYLIAQ
                     LKCFTQEKTSFKKRQVFLISRSLSIILANLLFYVVVIVWKANGFVYTSAMIFILIGGI
                     NFLSVTGTYIGLTILLYTAIAHQFWYMTKLTLTHCWLIIGLIWLLSAASSICIGLWGA
                     TLFYPESAPISCSFEKCQQPLAIIIVVGLSLCYGTVIGLYIAMMARLHRLVRKSSMIQ
                     ARSNSNSMTAMKRLGLNMITFAIGSIPILIVCIVALANLRELSSLGEGSKGPCKTFLH
                     SSLFVEVEILASVAAIVWLIAMIIDPVINILADKKMRGLFAKQINRVSQRIRRASTKS
                     TISTDDQ"
     gene            complement(14614. .15740)
                     /gene="ZK418.8"
     CDS             complement(join(14614. .14915,14961. .15184,15286. .15344,
                     15391. .15491,15543. .15628,15673. .15740))
                     /gene="ZK418.8"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50692.1"
                     /db_xref="PID:g470376"
                     /db_xref="GI:470376"
                     /translation="MGKLTNCLLAAMIASPFVIKYLRYRRLSAFVRLKPQLDDTDSEN
                     EADGILADASKDYFIRIPAFAVGAVVGKRGMTLRRLMYETQTECVLRSYTPEELDVAS
                     NEVSADAEWTEQFQNEDWFGEESTDTRRKITYLLVRATDEQCVKLVRLRIAEIIDNVQ
                     KPWKTAEFEVETQMVGYLVGRGGRHIKTLRDKFSVNIQISDPIPDDPTRSLVTVRGRD
                     LTALKEVEVVMKETVLPRNYNCARNRNYDEDTFLDNGIYDDVDEEIKDLKKVTELLAQ
                     DFF"
     gene            complement(16137. .20735)
                     /gene="ZK418.9"
     CDS             complement(join(16137. .16263,16312. .16790,16846. .17078,
                     17128. .17434,17482. .17807,17855. .17962,18265. .18328,
                     19466. .19632,20240. .20735))
                     /gene="ZK418.9"
                     /note="similar to yeast protein HX and E. coli PNP; coded
```

```
                     for by C. elegans cDNA cm2d8"
                     /codon_start=1
                     /protein_id="AAA50693.1"
                     /db_xref="PID:g470377"
                     /db_xref="GI:470377"
                     /translation="MDLAGLINNLSSLNINRNEYFMDYYAKEKSSVDNAVISTINFIK
                     KPPYLVFSMEEWTSINQLSTIDFIKKFDFVNQFSCYSDKSKFIHGVRFKSAIFSTSMR
                     VFNEKKSYMHFPGNVDIFPEHSSQLFNPHFLNKIRSQLIGKLIELDITKNESLLLSAL
                     FLCSAVHPDISTEGKNLLYKYQQYYSAALINHCCLTNQQNAPSRYSELLSLYQIIEET
                     HQKIIGADGGIGTLKRPLDSEILDGDLIPTKKSSEVGDLNMGDSDKITDIYPVPEKVV
                     GLVIGKGGSEIRLIQQTSGCRVQMDPDHQSVNGFRNCTIEGPPDQVAVARQMITQVIN
                     RNQTGAQPGAAPGEVTEEMLIPADKIGLVIGKGGETIRIVQEQSGLRNCNVVQETTTA
                     TGQPKPLRMIGSPAAIETAKALVHNIMNNTQGNAPLLQRAPHQPSGQFGGGYGAQEAQ
                     AKGEVIVPRLSAGMIIGKGGEMIKRLAAETGTKIQFKPDTNPNSEDRIAVIMGTRDQI
                     YRATERITEIVNRAIKNNGAPQDRGSAGTVLPGQSIFYMHVPAGKCGLVIGKGGENIK
                     QIERETGATCGLAPAAEQKNEDEKVFEIKGSQLQIHHASHLVRIKVGEISPNTPVPPL
                     QGAGGGYQQQQAMFSAGTQNGGYQSTGGFIQQQQQPQFQHQQQQWAPQNGGGVQQRLP
                     TEMYQNPMQQPQASVIPQASADPAPAVNPATGAPDYSAQWAAYYRSIGLVEQAAMVES
                     QMRRNQSNTAPLGDVSGVPQSQQPYYPQ"
     gene            22052. .24077
                     /gene="ZK418.2"
     CDS             join(22052. .22174,22217. .22268,22318. .22557,22762. .22988,
                     23035. .23114,23159. .23321,23409. .23510,23558. .23713,
                     23758. .23975,24023. .24077)
                     /gene="ZK418.2"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAA50686.1"
                     /db_xref="PID:g470370"

  Query Match             46.0%;  Score 439.2;  DB 36;  Length 32981;
  Best Local Similarity   74.4%;  Pred. No. 2.1e-93;
  Matches  698;  Conservative    0;  Mismatches    3;  Indels  237;  Gaps    2;

Qy     254 attatactcgtgctgatgaagagcccgggcgccaaggacgtccaccaggtcgccctcgta 313
           |||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2562 ATTTCAGTCGTGCTGATGAAGAGCCCGGGCGCCAAGGACGTCCACCAGGTCGCCCTCGTA 2503

Qy     314 agatgccgcgtcacgaatcttcaacttcacttatggaatcaccacgcaagactatgactc 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2502 AGATGCCGCGTCACGAATCTTCAACTTCACTTATGGAATCACCACGCAAGACTATGACTC 2443

Qy     374 gtgattctaaaattatgtttgaattgcgtggaaaaccattcgaaatgatagctggacgtt 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2442 GTGATTCTAAAATTATGTTTGAATTGCGTGGAAAACCATTCGAAATGATAGCTGGACGTT 2383

Qy     434 ttgaagaagaatattcacttggtagagcatgggttaaaggacacatgaataatgaatatg 493
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2382 TTGAAGAAGAATATTCACTTGGTAGAGCATGGGTTAAAGGACACATGAATAATGAATATG 2323

Qy     494 aaccaataaaagctcaaaggacagactatgcaccgaatctggctgtt------------ 541
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    2322 AACCAATAAAAGCTCAAAGGACAGACTATGCACCGAATCTGGCTGTTGTGAGTGGGAAAG 2263

Qy     541 ----------------------------------gattatcttgcatgtcgcgagattca 566
                                               ||||||||||||||||||||||||||
Db    2262 ATTAAATGAAAAATGTTATAATTATTTCCTTCAGGATTATCTTGCATGTCGCGAGATTCA 2203

Qy     567 tcgaatgccacgtccagataaatcaattcctgagctgccaattgttccatctagaatcga 626
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2202 TCGAATGCCACGTCCAGATAAATCAATTCCTGAGCTGCCAATTGTTCCATCTAGAATCGA 2143

Qy     627 tgaattcgacgctacagtcgatccaagatatgaaacagatttgaaaaatgaatacattcg 686
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2142 TGAATTCGACGCTACAGTCGATCCAAGATATGAAACAGATTTGAAAAATGAATACATTCG 2083

Qy     687 tcattggaaacaagtcaaaaaag---------------------------------- 710
           |||||||||||||||||||||||
Db    2082 TCATTGGAAACAAGTCAAAAAAGGGTAACACTTCTTCTGCCATTGAAAATTTCCGTAAAA 2023

Qy     710 ---------------------------------------------------------- 710

Db    2022 ATTTATTTCTTAAAGTATTACGGTAACTGTGCTCTATGTTAGAATTTCATTTGTCGAATT 1963

Qy     710 ---------------------------------------------------------- 710

Db    1962 TTGTATCAAACTATAAAATAAAAAATATTGTACTAGCAATCTGTGATCACTACTATTTCT 1903

Qy     710 --------------------------------gttggtgtgctcatcaacgtcgtcgga 736
                                            ||||||||||||||||||||||||||||
Db    1902 TCAAAGTCAGAGAATCCAAAAAAAAAAATTTCAGTTGGTGTGCTCATCAACGTCGTCGGA 1843

Qy     737 ctgctccccatgcaagaagcatagcattaatcaacaaaatctaccagcctggagagtcga 796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1842 CTGCTCCCCATGCAAGAAGCATAGCATTAATCAACAAAATCTACCAGCCTGGAGAGTCGA 1783

Qy     797 aaactgtcgagcaagcacttggtcttatttaaatattctaacatgtaatttcaatttatc 856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1782 AAACTGTCGAGCAAGCACTTGGTCTTATTTAAATATTCTAACATGTAATTTCAATTTATC 1723

Qy     857 tcttactttctgatcttgctatcacatgtctcttatttcaaaaatctcactttaaaattc 916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1722 TCTTACTTTCTGATCTTGCTATCACATGTCTCTTATTTCAAAAATCTCACTTTAAAATTC 1663

Qy     917 atataaataatgggtttattcaaatacatcatcttgac 954
           |||||||||||||||||||||||||||||||||||||
Db    1662 ATATAAATAATGGGTTTATTCAAATACATCATCTTGAC 1625


RESULT   2
AC006893
LOCUS       AC006893   260699 bp    DNA             HTG       26-FEB-1999
DEFINITION  Caenorhabditis elegans clone Y71F9, WORKING DRAFT SEQUENCE, 43
            unordered pieces.
ACCESSION   AC006893
NID         g4309906
VERSION     AC006893.2  GI:4309906
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 260699)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Caehorhabditis elegans clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 260699)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced gi:4263455.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3303: contig of 3303 bp in length
            *     3304     3312: gap of unknown length
            *     3313     6195: contig of 2883 bp in length
            *     6196     6204: gap of unknown length
            *     6205     8705: contig of 2501 bp in length
            *     8706     8714: gap of unknown length
            *     8715    27303: contig of 18589 bp in length
            *    27304    27312: gap of unknown length
            *    27313    29898: contig of 2586 bp in length
            *    29899    29907: gap of unknown length
            *    29908    32748: contig of 2841 bp in length
            *    32749    32757: gap of unknown length
            *    32758    42435: contig of 9678 bp in length
            *    42436    42444: gap of unknown length
            *    42445    45447: contig of 3003 bp in length
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 26, 1999, 10:23:43 ; Search time 90.19 Seconds
(without alignments)
1989.990 Million cell updates/sec

Title: US-09-087-136-2
Perfect score: 954
Sequence: 1 atgtcagaaatagatccact.....ttcaaatacatcatcttgac 954

Scoring table: IDENTITY_NUC

Searched: 240622 seqs, 94065609 residues

Database : N_Geneseq_34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	36.8	3.9	110000	1	V21209_04	Continuation (5 of
c	2	36.8	3.9	110000	1	V21209_13	Continuation (14 o
	3	36.8	3.9	110000	1	V21209_15	Continuation (16 o
c	4	36.6	3.8	5300	1	T90447	Thermoanaerobacter
c	5	35	3.7	110000	1	V21209_06	Continuation (7 of
	6	34.6	3.6	1576	1	Q40415	cDNA clone of pTOM
	7	34.4	3.6	4507	1	Q83525	SABP gene. New ery
	8	34.4	3.6	4507	1	T72888	Sialic acid bindin
c	9	34	3.6	110000	1	V21209_02	Continuation (3 of
	10	33.8	3.5	1463	1	V30584	Clostridium botuli
	11	33.8	3.5	1472	1	V30585	Clostridium botuli
	12	33.6	3.5	110000	1	V21209_06	Continuation (7 of
c	13	33.4	3.5	2872	1	Q25449	Cholesterol 7-alph
	14	33.2	3.5	4526	1	Q49780	Prosystemin. Trans
c	15	33	3.5	3265	1	N92188	TA29 gene insert i
	16	32.8	3.4	977	1	T73614	Werner's syndrome
	17	32.8	3.4	3580	1	V52342	Streptococcus pneu
	18	32.8	3.4	12666	1	V52270	Streptococcus pneu
c	19	32.8	3.4	58407	1	V21210	Methanococcus jann
	20	32.6	3.4	5510	1	Q98504	Human dystroglycan
c	21	32.6	3.4	705	1	T72057	Sequence flanking
c	22	32.6	3.4	705	1	T43937	Sequence flanking
	23	32.6	3.4	5510	1	T97317	Human dystrophin-a
	24	32.6	3.4	110000	1	V21209_12	Continuation (13 o
	25	32.6	3.4	110000	1	V21209_13	Continuation (14 o
	26	32.4	3.4	110000	1	V21209_10	Continuation (11 o
	27	32.2	3.4	1173	1	Q06504	CbEPV spheroidin g
c	28	32.2	3.4	110000	1	V21209_15	Continuation (16 o
c	29	32	3.4	5056	1	N90489	cDNA of human reti
c	30	32	3.4	2994	1	Q04713	Cancer supressing
c	31	32	3.4	4597	1	N81369	Human retinoblasto
c	32	32	3.4	4579	1	N81261	Probe for retinobl
c	33	32	3.4	2994	1	Q41545	Retinoblastoma gen
c	34	32	3.4	3789	1	Q44362	Sequence of bovine
c	35	32	3.4	4597	1	Q70536	Human retinoblasto
c	36	32	3.4	3233	1	Q72690	Retinoblastoma 94k
c	37	32	3.4	2995	1	Q86398	Human Rb110 cDNA.
c	38	32	3.4	2995	1	Q90059	Retinoblastoma tum
	39	32	3.4	110000	1	T58840_0	Mycoplasma genital
	40	32	3.4	110000	1	T58840_1	Continuation (2 of
	41	32	3.4	110000	1	V21209_07	Continuation (8 of
c	42	32	3.4	3461	1	V58447	Modified retinobla
c	43	32	3.4	3347	1	V58448	Modified retinobla
c	44	32	3.4	3161	1	V58449	Modified retinobla
c	45	32	3.4	3377	1	V58450	Modified retinobla

ALIGNMENTS

RESULT 1
V21209_04/c
Continuation (5 of 17) of V21209 from base 400001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 3.9%; Score 36.8; DB 1; Length 110000;
Best Local Similarity 54.4%; Pred. No. 3.5;
Matches 74; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy    815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttg 874
          ||  | ||| | ||||| ||   || |  || | ||| ||   ||| ||| |  | ||
Db  42462 TTATTATTAATAAAATAAACTTTAATTTTGTTACTATTAATGAGTTATTTTGTTTTTTTC 42403

Qy    875 ctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttta 934
           |||    | |  |||||||| ||  |       |||  |||    || |||  |   |||
Db  42402 ATATTTTTTATTACTTATTTAAATTAAAGACATTTTTAATAATGATTAATTTTTCTTTTT 42343

Qy    935 ttcaaatacatcatct 950
          |   |||| |  || ||
Db  42342 TAGTAATAAACTATTT 42327
```

RESULT 2
V21209_13/c
Continuation (14 of 17) of V21209 from base 1300001 (Methanococcus jannaschii circula
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 3.9%; Score 36.8; DB 1; Length 110000;

```
  Best Local Similarity   54.4%;  Pred. No. 3.5;
  Matches   74;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy     815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttg 874
           | ||| | || || || ||  | |      ||| |||| |||| |||   ||| || ||
Db   36481 TGGGTTTAATCTATATTTTAAAGCCCTCTTTTTTAATTCATCTTTTATAATCTTATTTTT 36422

Qy     875 ctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttta 934
             |   | || ||   | | |||        |  |   || |  | |||| ||| |  |
Db   36421 TTCAAATATTTCCTCTTTCTCATCCTCTGTAACTCCCAAATGTTTTAAAAAATTGTCTCT 36362

Qy     935 ttcaaatacatcatct 950
           ||||| ||| ||||
Db   36361 GACAAATTCATTATCT 36346


RESULT   3
V21209_15
Continuation (16 of 17) of V21209 from base 1500001 (Methanococcus jannaschii circular c
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP      Fragment Name       Begin      End
WP      V21209_00               1    110000
WP      V21209_01          100001    210000
WP      V21209_02          200001    310000
WP      V21209_03          300001    410000
WP      V21209_04          400001    510000
WP      V21209_05          500001    610000
WP      V21209_06          600001    710000
WP      V21209_07          700001    810000
WP      V21209_08          800001    910000
WP      V21209_09          900001   1010000
WP      V21209_10         1000001   1110000
WP      V21209_11         1100001   1210000
WP      V21209_12         1200001   1310000
WP      V21209_13         1300001   1410000
WP      V21209_14         1400001   1510000
WP      V21209_15         1500001   1610000
WP      V21209_16         1600001   1664976


  Query Match             3.9%;  Score 36.8;  DB 1;  Length 110000;
  Best Local Similarity   55.5%;  Pred. No. 3.5;
  Matches   71;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     821 ttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttgctatca 880
           || || || |  |||||  ||| |||| || ||| |||||    |  | | |   | |
Db   18390 TTTTTGAAGTTGTCTAAGTTGTTATTTAAAATTACCTCTTCAAATAGGCTTTCAATCTTG 18449

Qy     881 catgtctcttatttcaaaaatctcactttaaaattcatataaataatgggtttattcaaa 940
             | | |   |    | ||     || ||||| ||||   |  ||| |||||||| |||
Db   18450 ATATTTTTTGTTAAATATAAATATTCTCTAAAAGTCATTGGATAAATAGGTTTATGCAAT 18509

Qy     941 tacatcat 948
             | || |
Db   18510 GCCCTCCT 18517


RESULT   4
T90447/c
ID   T90447 standard; DNA; 5300 BP.
AC   T90447;
DT   27-JAN-1998  (first entry)
DE   Thermoanaerobacter thermohydrosulfuricus DNA polymerase gene.
KW   DNA polymerase; exonuclease; strand displacement amplification;
KW   Thermoanaerobacter thermohydrosulfuricans; reverse transcription;
KW   cDNA; ss.
OS   Thermoanaerobacter thermohydrosulfuricus.
FH   Key             Location/Qualifiers
FT   CDS             1056. .3674
FT                   /*tag= a
FN   WO9721821-A1.
PD   19-JUN-1997.
PF   13-DEC-1996; U20225.
PR   15-DEC-1995; US-008688.
PA   (AMSH ) AMERSHAM LIFE SCI INC.
PI   Davis M,  Mamone JA,  Sha D;
DR   WPI; 97-332792/30.
DR   P-PSDB; W26605.
PT   New DNA polymerase from Thermoanaerobacter thermohydrosulfuricus -
PT   useful for strand displacement amplification, reverse transcription
PT   and sequencing
PS   Example 1; Page 48-49; 92pp; English.
CC   This DNA sequence encodes a novel thermostable DNA polymerase
CC   (W26605) of the thermophilic anaerobe Thermoanaerobacter
CC   thermohydrosulfuricus.  It was isolated from a subgenomic library
CC   using a DNA polymerase I-specific probe that had itself been
CC   generated using primers (see T90452-53) based on conserved regions
CC   of DNA polymerase I family members.  A host cell comprising a
CC   vector containing an isolated nucleic acid encoding DNA polymerase
CC   is claimed.  The DNA polymerase can be used in claimed methods: (a)
CC   for strand displacement amplification (SDA) when used with a
CC   thermostable restriction enzyme; (b) for production of cDNA when
CC   used in conjunction with polymerase and deoxyribonucleotides; and
CC   (c) to provide reverse transcription activity in RT-PCR.  Mutagenic
CC   methods have been used to generate exonuclease-deficient DNA
CC   polymerases (see W26606-07) suited for sequencing, SDA and cDNA
CC   preparation.
SQ   Sequence   5300 BP;    2136 A;    587 C;    1036 G;    1541 T;


  Query Match             3.8%;  Score 36.6;  DB 1;  Length 5300;
  Best Local Similarity   58.9%;  Pred. No. 1.1;
  Matches   63;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     814 cttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatctt 873
           |||| |  ||  |||  || ||    ||||||  ||  |        | |||   | | ||
Db     599 CTTGCTTGTAAGTCATAATACTGGTCTGTAATGTCTTTGCTGTCAATCCTTAAAGCTGTT 540

Qy     874 gctatcacatgtctcttatttcaaaaatctcactttaaaattcatat 920
             || |  ||  ||| | |||| | ||||| |||||| ||||| || |
Db     539 ACTTTGCCAAATCTTTAATTTGAGAAATCACACTTTCAAATTTATCT 493


RESULT   5
V21209_06/c
Continuation (7 of 17) of V21209 from base 600001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP      Fragment Name       Begin      End
WP      V21209_00               1    110000
WP      V21209_01          100001    210000
WP      V21209_02          200001    310000
WP      V21209_03          300001    410000
WP      V21209_04          400001    510000
WP      V21209_05          500001    610000
WP      V21209_06          600001    710000
WP      V21209_07          700001    810000
WP      V21209_08          800001    910000
WP      V21209_09          900001   1010000
WP      V21209_10         1000001   1110000
WP      V21209_11         1100001   1210000
WP      V21209_12         1200001   1310000
WP      V21209_13         1300001   1410000
WP      V21209_14         1400001   1510000
WP      V21209_15         1500001   1610000
WP      V21209_16         1600001   1664976


  Query Match             3.7%;  Score 35;  DB 1;  Length 110000;
  Best Local Similarity   54.2%;  Pred. No. 11;
  Matches   71;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     803 tcgagcaagcacttggtcttatttaaatattctaacatgtaatttcaatttatctcttac 862
           || |  ||   | | |||| ||| | | |  | ||     | || | ||||||  ||||
Db   51285 TCTACTAACTTCATCCTCTTCTTTTAGTTTCATTACTAAATAGTTTATTTTATCCATTAC 51226

Qy     863 tttctgatcttgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataa 922
```

```
Db  51225 TACCTTGTTTAAATCCAATATCTCATTAATTTCAGATTTTTTCCTCTCAATTTCATTTTT 51166

Qy    923 ataatgggttt 933
Db  51165 AATTTCTTTTT 51155
```

RESULT 6
Q40415
ID Q40415 standard; cDNA; 1576 BP.
AC Q40415;
DT 05-AUG-1993 (first entry)
DE cDNA clone of pTOMUC82.1.
KW Clone; pTOMUC82.1; histidine decarboxylase-like; protein; HDC-like;
KW probe; isolation; developmental; regulatation; promoter; texture;
KW lambdaUC82-3.3; tissue-specific; modified phenotype; fruit; flavour;
KW solid content; processing quality;ss.
OS Lycopersicon esculentum.
PN WO9307257-A.
PD 15-APR-1993.
PF 02-OCT-1992; U08425.
PR 04-OCT-1991; US-770080.
PA (SMAR-) SMART PLANTS INT INC.
PI Dickinson CD, Elliott KJ, Fitzmaurice LC, Holtz GC;
PI Mirkov TE;
DR WPI; 93-134448/16.
PT Tissue-specific, developmentally regulated transcriptional
PT sequences - modify phenotype(s) of fruit, esp. tomatoes, for
PT herbicide tolerance, improved flavour, etc.
PS Claim 12; Page 21-22; 30pp; English.
CC This sequence represents a cDNA clone of pTOMUC82.1. This sequence
CC encodes a histidine decarboxylase-like protein (HDC-like). Portions
CC of this sequence can be used as probes to isolate genes which may
CC have tissue-specific, developmentally regulatable promoters eg.
CC lambdaUC82-3.3 (see also Q40416). Tissue-specific promoters may be
CC used with exogenous DNA to provide plants having tissues with a
CC modified phenotypic property. The resulting fruit may have improved
CC solids content, flavour, texture, processing qualities etc.
SQ Sequence 1576 BP; 529 A; 252 C; 291 G; 504 T;

Query Match 3.6%; Score 34.6; DB 1; Length 1576;
Best Local Similarity 46.8%; Pred. No. 2.3;
Matches 109; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```
Qy    695 aacaagtcaaaaaaggttggtgtgctcatcaacgtcgtcggactgctccccatgcaagaa 754
Db    656 AAACACTTGAAAATTGTGGTTATTCAAATGACAATTATTATATCCATTGCGATGCAGCAT 715

Qy    755 gcatagcattaatcaacaaaatctaccagcctggagagtcgaaaactgtcgagcaagcac 814
Db    716 TATGTGGGCTAATTCTCCCATTTATCAAACATGCAAAAAAAATTACCTTCAAGAAGCCAA 775

Qy    815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttacttttctgatcttg 874
Db    776 TTGGTAGTATTTCAATTTCAGGGCACAAATTCTTGGGATGTCCAATGCCTTGTGGCATTC 835

Qy    875 ctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataat 927
Db    836 AGATAACAAGGAAAACTTATGTTAGTACCCACTCAAAAATTGAGTATATTAAT 888
```

RESULT 7
Q83525
ID Q83525 standard; DNA; 4507 BP.
AC Q83525;
DT 22-SEP-1995 (first entry)
DE SABP gene.
KW SABP; sialic acid binding protein; binding domain; merozoite;
KW malaria; therapy; vaccine; ss.
OS Plasmodium falciparum.
PN WO9507353-A.
PD 16-MAR-1995.
PF 07-SEP-1994; U10230.
PR 10-SEP-1993; US-119677.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellems TE;
DR WPI; 95-123427/16.
DR P-PSDB; R70232.
PT New erythrocyte binding domain polypeptide(s) - isolated from
PT Plasmodium binding proteins, used in diagnosis, treatment and
PT prevention of malaria
PS Disclosure; Page 39-41; 81pp; English.
CC Sequences from the SABP gene (given in Q83525) were PCR amplified,
CC expressed on the surface of COS cells and tested for erythrocyte
CC binding to identify the binding domain polypeptide. A prefd. SABP
CC binding domain comprises residues 1 to about 616 of the SABP protein
CC (R70232). Recombinant binding domain was expressed in E. coli,
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-
CC infected cells. It provides protection against P. falciparum.
SQ Sequence 4507 BP; 1955 A; 458 C; 796 G; 1298 T;

Query Match 3.6%; Score 34.4; DB 1; Length 4507;
Best Local Similarity 45.2%; Pred. No. 4.1;
Matches 168; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

```
Qy    371 ctcgtgattctaaaattatgtttgaattgcgtggaaaaccattcgaaatgatagctggac 430
Db   2571 CTAATGATAATATAAATGTTACTGAACAAGGGGATAATATTTCCGGGGTGAATTCTAAAC 2630

Qy    431 gttttgaagaagaatattcacttggtagagcatgggttaaaggacacatgaataatgaat 490
Db   2631 CTTTATCTGATGATGTACGTCCAGATAAAAAGGAATTAGAAGATCAAAATAGTGATGAAT 2690

Qy    491 atgaaccaataaaagctcaaaggacagactatgcaccgaatctggctgttgattatcttg 550
Db   2691 CGGAAGAAACTGTAGTAAATCATATATCAAAAAGTCCATCTATAAATAATGGAGATGATT 2750

Qy    551 catgtcgcgagattcatcgaatgccacgtccagataaatcaattcctgagctgccaattg 610
Db   2751 CAGGCAGTGGAAGTGCAACAGTGAGTGAATCTAGTAGTTCAAATACTGGATTGTCTATTG 2810

Qy    611 ttccatctagaatcgatga---attcgacgctacagtcgatccaagatatgaaacagatt 667
Db   2811 ATGATGATAGAAATGGTGATACATTTGTTCGAACACAAGATACAGCAAATACTGAAGATG 2870

Qy    668 tgaaaaatgaatacattcgtcattggaaacaagtcaaaaaaggttggtgtgctcatcaac 727
Db   2871 TTATTAGAAAAGAAAATGCTGACAAGGATGAAGATGAAAAAGGCGCAGATGAAGAAAGAC 2930

Qy    728 gtcgtcggactg 739
Db   2931 ATAGTACTTCTG 2942
```

RESULT 8
T72888
ID T72888 standard; DNA; 4507 BP.
AC T72888;
DT 11-SEP-1997 (first entry)
DE Sialic acid binding protein coding sequence.
KW DBL gene family; SABP; sialic acid binding protein; merozoite; malaria;
KW Duffy antigen binding protein; DABP; erythrocyte; var-1; var-2; var-3;
KW var-7; vaccine; therapy; immune response; Plasmodium; ss.
OS Plasmodium falciparum.
FH Key Location/Qualifiers
FT cds 56. .4363
FT /*tag= a
FT /product= Duffy antigen binding protein
PN WO9640766-A2.
PD 19-DEC-1996.
PF 07-JUN-1996; U09508.
PR 07-JUN-1995; US-487826.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellems TE;
DR WPI; 97-052231/05.
DR P-PSDB; W22477.
PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins
PS Example 1; Page 35-36; 96pp; English.
CC This sequence represents the full length coding sequence for the sialic
CC acid binding protein (SABP). SABP and the Duffy antigen binding protein
CC (DABP) are soluble proteins that appear in the culture supernatant after
CC infected erythrocytes release merozoites. DABP and SABP mediate the
CC binding of merozoites and schizonts to the erythrocyte surface. These
CC proteins are necessary for erythrocyte invasion by the parasite. This
CC sequence can be used in the compositions of the invention. The
CC compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 4507 BP; 1955 A; 458 C; 796 G; 1298 T;

Query Match 3.6%; Score 34.4; DB 1; Length 4507;
Best Local Similarity 45.2%; Pred. No. 4.1;
Matches 168; Conservative 0; Mismatches 201; Indels 3; Gaps 1;

```
Qy     371 ctcgtgattctaaaattatgtttgaattgcgtggaaaaccattcgaaatgatagctggac 430
           ||  ||||  || || | |   ||||    | |  ||    | ||   |||   ||  ||
Db    2571 CTAATGATAATATAAATGTTACTGAACAAGGGGATAATATTTCCGGGGTGAATTCTAAAC 2630

Qy     431 gttttgaagaagaatattcacttggtagagcatgggttaaaggacacatgaataatgaat 490
           |||    || ||       |  | || |       |  |||  || |  | | ||||||
Db    2631 CTTTATCTGATGATGTACGTCCAGATAAAAAGGAATTAGAAGATCAAAATAGTGATGAAT 2690

Qy     491 atgaaccaataaaagctcaaaggacagactatgcaccgaatctggctgttgattatcttg 550
            |||  ||   ||   |    | |    |    ||   | |  |  ||   ||  |
Db    2691 CGGAAGAAACTGTAGTAAATCATATATCAAAAAGTCCATCTATAAATAATGGAGATGATT 2750

Qy     551 catgtcgcgagattcatcgaatgccacgtccagataaatcaattcctgagctgccaattg 610
           || |  | |  | |     | ||       |   ||  |||| | |||   || | ||||
Db    2751 CAGGCAGTGGAAGTGCAACAGTGAGTGAATCTAGTAGTTCAAATACTGGATTGTCTATTG 2810

Qy     611 ttccatctagaatcgatga---attcgacgctacagtcgatccaagatatgaaacagatt 667
           |     |||||  | |||   ||| |     |||   ||| ||  | ||     ||||
Db    2811 ATGATGATAGAAATGGTGATACATTTGTTCGAACACAAGATACAGCAAATACTGAAGATG 2870

Qy     668 tgaaaaatgaatacattcgtcattggaaacaagtcaaaaaaggttggtgtgctcatcaac 727
           | |  |   || | | |  | |   | |  |||   |||||||      ||   |  ||
Db    2871 TTATTAGAAAAGAAAATGCTGACAAGGATGAAGATGAAAAAGGCGCAGATGAAGAAAGAC 2930

Qy     728 gtcgtcggactg 739
           | ||    |||
Db    2931 ATAGTACTTCTG 2942
```

RESULT 9
V21209_02/c
Continuation (3 of 17) of V21209 from base 200001 (Methanococcus jannaschii circular chr
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976

Query Match 3.6%; Score 34; DB 1; Length 110000;
Best Local Similarity 54.9%; Pred. No. 21;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

```
Qy     814 cttggtcttatttaaatattctaacatgtaatttcaatttatctcttacttttctgatctt 873
           |||| | ||| ||  || || |  |    |||| |  ||  | |  | | || |  | ||
Db   78437 CTTGTTTTTACTTTTATTTTTTCTCCATCAATTACTTTTGTTATTGTTCCTTTTCTTTTT 78378

Qy     874 gctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttt 933
             ||  ||| | ||    |||  ||||| ||  ||||||||   | | |  |     |||
Db   78377 TTCATGGCATTTTTTACCTTTTTAAAATTTCCTTTTAAAATGTTTTTTATAATCCATTTT 78318

Qy     934 at 935
           ||
Db   78317 AT 78316
```

RESULT 10
V30584
ID V30584 standard; DNA; 1463 BP.
AC V30584;
DT 07-DEC-1998 (first entry)
DE Clostridium botulinum toxin E fragment C gene in pETHisb.
KW Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW botulism; BotE; ds.
OS Clostridium botulinum serotype E strain Belgua.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 108. .1463
FT /*tag= a
PN WO9808540-A1.
PD 05-MAR-1998.
PF 28-AUG-1997; U15394.
PR 28-AUG-1996; US-704159.
PA (OPHI-) OPHIDIAN PHARM INC.
PI Thalley BS, Williams JA;
DR WPI; 98-230234/20.
DR P-PSDB; W68395.
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
PS Example 41; Page 324-325; 428pp; English.
CC This is the DNA sequence of the Clostridium botulinum serotype E
CC (Beluga strain) neurotoxin fragment C gene contained in plasmid
CC pETHisb. The encoded BotE fragment C polypeptide (see W68395) has a
CC His-tagged N-terminal extension. The vector was used to express
CC native (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins, especially type B and type E toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred
CC hosts for production of recombinant proteins are E. coli, insect
CC cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are
CC useful in the treatment of humans and animals at risk of
CC intoxication with clostridial toxin.
SQ Sequence 1463 BP; 598 A; 150 C; 217 G; 498 T;

Query Match 3.5%; Score 33.8; DB 1; Length 1463;
Best Local Similarity 52.5%; Pred. No. 3.7;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
Qy    813 acttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatct 872
          ||||  | ||| |  ||  ||   | ||  ||  |||||||     |||  |  |  ||
Db    719 ACTTTATATTAATGGAAATTTAATAGATAAAAAATCAATTTTAAATTTAGGTAATATTCA 778

Qy    873 tgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggtt 932
          || ||       | | |||||| ||| |  | | | ||    |   | | ||| |||  |
Db    779 TGTTAGTGACAATATATTATTTAAAATAGTTAATTGTAGTTATACAAGATATATTGGTAT 838

Qy    933 tattcaaatacatcatcttga 953
          ||   |  |  ||  | ||||
Db    839 TAGATATTTTAATATTTTTGA 859


RESULT  11
V30585
ID   V30585 standard; DNA; 1472 BP.
AC   V30585;
DT   07-DEC-1998  (first entry)
DE   Clostridium botulinum toxin E fragment C gene in pETHisb.
KW   Antitoxin; vaccine; neurotoxin; toxin E; intoxication; immunogen;
KW   botulism; BotE; ds.
OS   Clostridium botulinum serotype E strain NCTC 11219.
OS   Synthetic.
FH   Key             Location/Qualifiers
FT   CDS             108. .1466
FT                   /*tag=  a
PN   WO9808540-A1.
PD   05-MAR-1998.
PF   28-AUG-1997; U15394.
PR   28-AUG-1996; US-704159.
PA   (OPHI-) OPHIDIAN PHARM INC.
PI   Thalley BS, Williams JA;
DR   WPI; 98-230234/20.
DR   P-PSDB; W68396.
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
PS   Example 41; Page 327-329; 428pp; English.
CC   This is the DNA sequence of the Clostridium botulinum serotype E
CC   (NCTC 11219) neurotoxin fragment C gene contained in plasmid
CC   pETHisb.  The encoded BotE fragment C polypeptide (see W68396) has a
CC   His-tagged N-terminal extension.  The vector was used to express
CC   native (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC   cells.  The invention relates to recombinant proteins derived from
CC   C. botulinum toxins, especially type B and type E toxins.  Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination.  Preferred
CC   hosts for production of recombinant proteins are E. coli, insect
CC   cells and yeast cells.  The recombinant toxins are used as
CC   immunogens for the production of vaccines and antitoxins that are
CC   useful in the treatment of humans and animals at risk of
CC   intoxication with clostridial toxin.
SQ   Sequence   1472 BP;    602 A;    151 C;    221 G;    498 T;


  Query Match             3.5%;  Score 33.8;  DB 1;  Length 1472;
  Best Local Similarity   52.5%;  Pred. No. 3.7;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy    813 acttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatct 872
          ||||  | ||| |  ||  ||   | ||  ||  |||||||     |||  |  |  ||
Db    719 ACTTTATATTAATGGAAATTTAATAGATCAAAAATCAATTTTAAATTTAGGTAATATTCA 778

Qy    873 tgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggtt 932
          || ||       | | |||||| ||| |  | | | ||    |   | | ||| |||  |
Db    779 TGTTAGTGACAATATATTATTTAAAATAGTTAATTGTAGTTATACAAGATATATTGGTAT 838

Qy    933 tattcaaatacatcatcttga 953
          ||   |  |  ||  | ||||
Db    839 TAGATATTTTAATATTTTTGA 859


RESULT  12
V21209_06
Continuation (7 of 17) of V21209 from base 600001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments  LOCUS V21209 Accession V21209
WP      Fragment Name      Begin     End
WP      V21209_00              1    110000
WP      V21209_01         100001    210000
WP      V21209_02         200001    310000
WP      V21209_03         300001    410000
WP      V21209_04         400001    510000
WP      V21209_05         500001    610000
WP      V21209_06         600001    710000
WP      V21209_07         700001    810000
WP      V21209_08         800001    910000
WP      V21209_09         900001   1010000
WP      V21209_10        1000001   1110000
WP      V21209_11        1100001   1210000
WP      V21209_12        1200001   1310000
WP      V21209_13        1300001   1410000
WP      V21209_14        1400001   1510000
WP      V21209_15        1500001   1610000
WP      V21209_16        1600001   1664976


  Query Match             3.5%;  Score 33.6;  DB 1;  Length 110000;
  Best Local Similarity   53.9%;  Pred. No. 27;
  Matches   69;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy    797 aaactgtcgagcaagcacttggtcttatttaaatattctaacatgtaatttcaatttatc 856
          ||| | || |   |    | | | |||||||    || |   |  |  |    |||||||
Db  33289 AAAGTTTCTAAGTAATCGTGGATTTTATTTAGGAGTTTTTCAAAATTCTCATCATTTATC 33348

Qy    857 tcttactttctgatcttgctatcacatgtctcttatttcaaaaatctcactttaaaattc 916
          ||||  | | | |  || | |||  |    |||| |  ||   |  | |||||||  |||
Db  33349 TCTTCATCTATAACTTTTCCATCCAACAAATCTTCATCTAATTTTAGCTCTTTAATCTTC 33408

Qy    917 atataaat 924
          |  ||||
Db  33409 TTTAAAAT 33416


RESULT  13
Q25449/c
ID   Q25449 standard; DNA; 2872 BP.
AC   Q25449;
DT   07-DEC-1992  (first entry)
DE   Cholesterol 7-alpha-hydrolase gene.
KW   C7H; human; 7-hydroxycholesterol; recombinant; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   cds             40. .504
FT                   /*tag=  a
PN   J04144680-A.
PD   19-MAY-1992.
PF   05-SEP-1990; 268587.
PR   05-SEP-1990; JP-268587.
PA   (TOXN ) Toyo jozo kk.
DR   WPI; 92-214120/26.
DR   P-PSDB; R24482.
PT   Cholesterol 7-alpha-hydroxylase enzyme - and DNA sequence etc., for
PT   efficient prodn. of enzyme
PS   Claim 1; Fig 1; 15pp; Japanese.
CC   The DNA sequence encodes human cholesterol 7-alpha-hydrolase which
CC   catalyses the formation of 7-alpha-hydroxycholesterol, NADP and
CC   water from cholesterol, reduced NADP and molecular state oxygen.
CC   The enzyme may be produced by recombinant techniques by culturing
CC   cells transformed with a plasmid contg. the DNA sequence and
CC   recovering the enzyme from the culture.  C7H may be produced
CC   efficiently by this method.
SQ   Sequence   2872 BP;    898 A;    532 C;    561 G;    881 T;


  Query Match             3.5%;  Score 33.4;  DB 1;  Length 2872;
```

```
  Best Local Similarity   48.2%;  Pred. No. 6.3;
  Matches   94;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     751 agaagcatagcattaatcaacaaaatctaccagcctggagagtcgaaaactgtcgagcaa 810
           ||| |||| ||| | |||  |     |  | | |  |||  | ||  |
Db    2851 AGATTCATACTTTTATTTCTGAAAGATGAATCATTTCAAAATTATCAAATGTTAGAATAT 2792

Qy     811 gcacttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgat 870
           ||| | ||  | |||||  | || | |||| |  ||||||  | | | |     |
Db    2791 TTTCTTAATATTTGTAAAATAAGCAAAAAGGTAAATATAATTTAAATATAAATACTACAC 2732

Qy     871 cttgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatggg 930
           |||||  || |    |  | | || ||  || ||||||    |    | |
Db    2731 AGAAATATCAATTGACATAGAGCTGATAACTCATACCTTAAAATATTTTATCAAAGCATT 2672

Qy     931 tttattcaaatacat 945
           || | || |  |||
Db    2671 GTTGTACACAGGCAT 2657


RESULT  14
Q49780
ID   Q49780 standard; DNA; 4526 BP.
AC   Q49780;
DT   26-APR-1994  (first entry)
DE   Prosystemin.
KW   Systemin; tomato; plant; resistance; insect; wounding;
KW   hormone; defense; predator; herbivore; bacteria; virus; ds.
OS   Lycopersicon esculentum (v. Castlemart).
PN   WO9319079-A.
PD   30-SEP-1993.
PF   18-MAR-1993; U02428.
PR   19-MAR-1992; US-855412.
PA   (UNIW ) UNIV WASHINGTON STATE RES FOUND.
PI   Mcgurl BF,  Pearce GL,  Ryan CA;
DR   WPI; 93-320672/40.
PT   Transgenic plants contg. systemin cDNA - show increased
PT   resistance to wounding by insects
PS   Claim 9; Page 38-41; 75pp; English.
CC   Systemin, an 18 amino acid peptide hormone (R41939) and first
CC   polypeptide hormone found in plants, induces expression of defense
CC   in plants, wounded mechanically or by predators including
CC   herbivores, insects, bacteria and viruses.  The precursor for systemin
CC   is encoded as a 200 amino acid prosystemin molecule (R41938) that has
CC   the systemin peptide sequence located near the C-terminus.  Both a cDNA
CC   and genomic DNA were cloned (Q49779-Q49780 respectively).  Transgenic
CC   plants constructed with antisense prosystemin DNA fail to mount a
CC   defensive response to wounding.  Transgenic plants constructed with
CC   increased copy number of prosystemin genes exhibit increased resistance
CC   to wounding.  Insect larva that feed on transgenic plants constructed
CC   with increased copy number of prosystemin genes exhibit decreased growth
CC   weight compared to larval that feed on wild type plants.
SQ   Sequence   4526 BP;    1621 A;    570 C;    530 G;    1805 T;


  Query Match             3.5%;  Score 33.2;  DB 1;  Length 4526;
  Best Local Similarity   57.8%;  Pred. No. 8.8;
  Matches   59;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy     830 tattctaacatgtaatttcaatttatctcttactttctgatcttgctatcacatgtctct 889
           |||| |    |    |||||  |||| | |   |  || |  ||  ||| | |  | | |
Db    2538 TATTTTCTTCTTATTTTTCATATTATTTTTATTTAACTAAATTTAGTATGAAACCTTTTT 2597

Qy     890 tatttcaaaaatctcactttaaaattcatataaataatgggt 931
           | ||| |||||  |||  |||||    |||| | |  |||||
Db    2598 TTTTTTAAAAAAATCATCTTAAATAAAATATTATTTTTGGGT 2639


RESULT  15
N92188/c
ID   N92188 standard; DNA; 3265 BP.
AC   N92188;
DT   11-APR-1990  (first entry)
DE   TA29 gene insert in pTA29S3.
KW   TA29 promoter; pTA29S3.
OS   Nicotiana tabacum "Samsun".
FH   Key             Location/Qualifiers
FT   tata_signal      1445. .1451
FT                    /*tag=  a
FT   cds              1526. .4991
FT                    /*tag=  b
FT   cds              1940. .2296
FT                    /*tag=  c
PN   WO8910396-A.
PD   02-NOV-1989.
PF   27-APR-1989; E00495.
PR   28-APR-1988; GB-010120.
PA   (PLAN-) Plant Genetic Systems NV.
PI   Mariani C, Leemans J, De Greef W, De Beuckeleer M;
DR   WPI; 89-339961/46.
DR   P-PSDB; P93313.
PT   Transformed plant cell with disrupted metabolism - from a product encoded
PT   by foreign male-sterility DNA inserted in the nuclear genome.
PS   Example 2; fig. 3A; 89pp; English.
CC   Plasmid pTA29S3 was constructed subcloning a ClaI fragment  contg. the
CC   TA29 gene from lambda TA29 into pLK31. The TA29 promoter is used in a
CC   vector for transforming plant cells. Tag b is the coding sequence of the
CC   TA29 gene; tag b is the cDNA insert from pBR329 (N92187).
SQ   Sequence   3265 BP;    1016 A;    581 C;    623 G;    1045 T;


  Query Match             3.5%;  Score 33;  DB 1;  Length 3265;
  Best Local Similarity   53.5%;  Pred. No. 8.6;
  Matches   69;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     824 tttaaatattctaacatgtaatttcaatttatctcttactttctgatcttgctatcacat 883
           ||| | | |  |||  |  |||| ||| | || |  |||||| ||  | |  | |   |
Db    3051 TTTGAGTGTAATAATGTTAAATTACAAATGATATGCTACTTTATGGCCATCTTGTTGAAA 2992

Qy     884 gtctcttatttcaaaaatctcactttaaaattcatataaataatgggtttattcaaatac 943
           |||||  |||  |  |  | | |   || |||    ||| ||  | | | |    | |
Db    2991 TAATCTTAAATCATATGTTGCTCCTACGAAATCAATCCAATCATCTGATCAATAGCTTTC 2932

Qy     944 atcatcttg 952
           | |||  ||
Db    2931 AGCATTATG 2923


Search completed: August 26, 1999, 11:25:06
Job time: 3683 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 26, 1999, 10:12:01 ; Search time 507.8 Seconds
(without alignments)
2776.189 Million cell updates/sec

Title: US-09-087-136-2
Perfect score: 954
Sequence: 1 atgtcagaaatagatccact.....ttcaaatacatcatcttgac 954

Scoring table: IDENTITY_NUC

Searched: 1944271 seqs, 738861917 residues

Database :
EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: em_est20:*
45: em_est21:*
46: em_est22:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1		354	37.1	360	24	D74918	D74918 CELK090A5F
2		354	37.1	360	24	D75085	D75085 CELK096B5F
3		334.4	35.1	360	24	D69562	D69562 CELK071DZF
4	c	314.4	33.0	339	24	D65944	D65944 CELK071DZR
5	c	300	31.4	300	35	C55654	C55654 C55654 Yuji
6	c	173.8	18.2	180	24	D72252	D72252 CELK096B5R
7		45.6	4.8	438	29	AA163826	AA163826 mr18g07.r
8	c	40.4	4.2	280	39	C90616	C90616 C90616 Dict
9	c	38.4	4.0	466	39	C84248	C84248 C84248 Dict
10		37.6	3.9	391	22	R52024	R52024 yg84h10.s1
11		37.2	3.9	478	20	M80104	M80104 wEST00641 M
12		37.2	3.9	652	34	AA550683	AA550683 1920m3 gm
13	c	37.2	3.9	342	43	AI250750	AI250750 qi36a05.x
14	c	37	3.9	459	40	C94326	C94326 C94326 Dict
15	c	36.8	3.9	571	34	AA518097	AA518097 vi23b01.r
16	c	36.8	3.9	419	39	C90267	C90267 C90267 Dict
17		36.6	3.8	404	29	AA158298	AA158298 zo53b03.r
18		36.4	3.8	463	27	C10394	C10394 C10394 Yuji
19	c	36.4	3.8	354	36	AA654946	AA654946 nt77g11.s
20	c	36.2	3.8	329	41	AI091286	AI091286 ow62c07.x
21		36	3.8	413	22	H02435	H02435 yj40g06.r1
22		36	3.8	413	22	R61290	R61290 yh07a01.s1
23		36	3.8	426	22	R70682	R70682 yi41d10.r1
24		36	3.8	500	22	R77183	R77183 yi65e03.r1
25	c	36	3.8	244	25	W18214	W18214 SWAMCA661SK
26		36	3.8	483	27	AA055260	AA055260 zf20e10.r
27		36	3.8	399	29	AA162388	AA162388 mt69a04.r
28		36	3.8	216	33	AA442294	AA442294 zv54c05.r
29		36	3.8	379	39	AA865198	AA865198 oh28c09.s
30	c	35.8	3.8	500	24	N43069	N43069 SW3ICA1116S
31	c	35.8	3.8	594	26	W51601	W51601 AS4SLC872T3
32	c	35.8	3.8	449	28	AA072493	AA072493 SW3ICA163
33	c	35.8	3.8	417	38	AA842075	AA842075 MBAFCF3A0
34	c	35.8	3.8	874	40	AA991085	AA991085 BSBmMFSZ0
35	c	35.8	3.8	537	40	C93331	C93331 C93331 Dict
36	c	35.8	3.8	549	41	AI052927	AI052927 SWBmL3SBH
37	c	35.6	3.7	465	21	T80527	T80527 yd08b04.r1
38	c	35.6	3.7	444	24	N43209	N43209 SW3ICA1131S
39	c	35.4	3.7	276	41	AI052863	AI052863 SWBmL3SBH
40		35.2	3.7	429	26	W62058	W62058 md86e01.r1
41	c	35.2	3.7	505	39	C84048	C84048 C84048 Dict
42		35	3.7	502	34	AA549856	AA549856 0929m3 gm
43	c	34.8	3.6	453	28	AA070345	AA070345 zm68h03.r
44	c	34.8	3.6	803	35	AA585652	AA585652 SW3D9CA39
45		34.8	3.6	476	40	C93091	C93091 C93091 Dict

ALIGNMENTS

RESULT 1
D74918
LOCUS D74918 360 bp mRNA EST 14-DEC-1995
DEFINITION CELK090A5F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk90a5 5', mRNA sequence.
ACCESSION D74918
NID g1120702
VERSION D74918.1 GI:1120702
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT
Contact: Yuji Kohara

Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddbj.nig.ac.j
High quality sequence stop: 370.

```
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
                     /db_xref="taxon:6239"
                     /clone="yk90a5"
                     /clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT      109 a     95 c     77 g     78 t      1 others
ORIGIN
```

Query Match 37.1%; Score 354; DB 24; Length 360;
Best Local Similarity 100.0%; Pred. No. 1e-82;
Matches 354; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgtcagaaatagatccacttgccgagttcttgcttccagaagacggagatcgaaatgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6 ATGTCAGAAATAGATCCACTTGCCGAGTTCTTGCTTCCAGAAGACGGAGATCGAAATGCT 65

Qy      61 cgtcaaaatgatccattgataagcggaggtccacttccattggaatcgccaagcagaaaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      66 CGTCAAAATGATCCATTGATAAGCGGAGGTCCACTTCCATTGGAATCGCCAAGCAGAAAA 125

Qy     121 ctcacatccctgttatcctatgatccgacagttccggagtcaccggatatgaaattcgcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126 CTCACATCCCTGTTATCCTATGATCCGACAGTTCCGGAGTCACCGGATATGAAATTCGCC 185

Qy     181 agaaaacgtctgggaaatctgctgacaaccataaaacatcacccatcggaaataattgga 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     186 AGAAAACGTCTGGGAAATCTGCTGACAACCATAAAACATCACCCATCGGAAATAATTGGA 245

Qy     241 gtactcccagaagattatactcgtgctgatgaagagcccgggcgccaaggacgtccacca 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     246 GTACTCCCAGAAGATTATACTCGTGCTGATGAAGAGCCCGGGCGCCAAGGACGTCCACCA 305

Qy     301 ggtcgccctcgtaagatgccgcgtcacgaatcttcaacttcacttatggaatca 354
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     306 GGTCGCCCTCGTAAGATGCCGCGTCACGAATCTTCAACTTCACTTATGGAATCA 359
```

```
RESULT   2
D75085
LOCUS       D75085        360 bp    mRNA            EST       14-DEC-1995
DEFINITION  CELK096B5F Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk96b5 5', mRNA sequence.
ACCESSION   D75085
NID         g1120869
VERSION     D75085.1  GI:1120869
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT
            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
```

Fax: 0559-75-6240
Email: ykohara@ddbj.nig.ac.j
High quality sequence stop: 282.

```
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
                     /db_xref="taxon:6239"
                     /clone="yk96b5"
                     /clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT      109 a     92 c     76 g     77 t      6 others
ORIGIN
```

Query Match 37.1%; Score 354; DB 24; Length 360;
Best Local Similarity 98.3%; Pred. No. 1e-82;
Matches 354; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy       9 aatagatccacttgccgagttcttgcttccagaagacggagatcgaaatgctcgtcaaaa 68
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AATAGATCCACTTGCCGAGTTCTTGCTTCCAGAAGACGGAGATCGAAATGCTCGTCAAAA 60

Qy      69 tgatccattgataagcggaggtccacttccattggaatcgccaagcagaaaactcacatc 128
           |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db      61 TGATCCATTGATAAGCGGAGGTCCACTTCCATTGGAATCGNCAAGNAGAAAACTCACATC 120

Qy     129 cctgttatcctatgatccgacagttccggagtcaccggatatgaaattcgccagaaaacg 188
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     121 CCTGTTATCCTATGATCCGACAGTTCCGGAGTCACCGGATATGAAATTCGNCAGAAAACG 180

Qy     189 tctgggaaatctgctgacaaccataaaacatcacccatcggaaataattggagtactccc 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TCTGGGAAATCTGCTGACAACCATAAAACATCACCCATCGGAAATAATTGGAGTACTCCC 240

Qy     249 agaagattatactcgtgctgatgaagagcccgggcgccaaggacgtccaccaggtcgccc 308
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     241 AGAAGATTATACTCGTGCTGATGAAGAGCCCGGGCGNCAAGGACGTCCACCAGGTCGCCC 300

Qy     309 tcgtaagatgccgcgtcacgaatcttcaacttcacttatggaatcaccacgcaagactat 368
           |||||||||||| |||||||||| |||||||||||||||||||||||||||||||||||
Db     301 TCGTAAGATGCCGNGTCACGAATCTNCAACTTCACTTATGGAATCACCACGCAAGACTAT 360
```

```
RESULT   3
D69562
LOCUS       D69562        360 bp    mRNA            EST       07-DEC-1995
DEFINITION  CELK071DZF Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk71d12 5', mRNA sequence.
ACCESSION   D69562
NID         g1105213
VERSION     D69562.1  GI:1105213
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:838186.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j
            Insert Length: 1251   Std Error: 0.00
```

High quality sequence stop: 321.

```
FEATURES             Location/Qualifiers
     source          1. .360
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
                     /db_xref="taxon:6239"
                     /clone="yk71d12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT      107 a     96 c     77 g     78 t      2 others
ORIGIN
```

```
  Query Match           35.1%;  Score 334.4;  DB 24;  Length 360;
  Best Local Similarity  97.2%;  Pred. No. 1.3e-77;
  Matches  350;  Conservative    0;  Mismatches    8;  Indels    2;  Gaps    1;

Qy       3 gtcagaaatagatccacttgccgagttcttgcttccagaagacggag--atcgaaatgct 60
           ||||||||||||||||||||||||||||||||||||||||||||||    |    ||||||||||
Db       1 GTCAGAAATAGATCCACTTGCCGAGTTCTTGCTTCCAGAAAGACGGAGTATCGAAATGCT 60

Qy      61 cgtcaaaatgatccattgataagcggaggtccacttccattggaatcgccaagcagaaaa 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CGTCAAAATGATCCATTGATAAGCGGAGGTCCACTTCCATTGGAATCGCCAAGCAGAAAA 120

Qy     121 ctcacatccctgttatcctatgatccgacagttccggagtcaccggatatgaaattcgcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CTCACATCCCTGTTATCCTATGATCCGACAGTTCCGGAGTCACCGGATATGAAATTCGCC 180

Qy     181 agaaaacgtctgggaaatctgctgacaaccataaaacatcacccatcggaaataattgga 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGAAAACGTCTGGGAAATCTGCTGACAACCATAAAACATCACCCATCGGAAATAATTGGA 240

Qy     241 gtactcccagaagattatactcgtgctgatgaagagcccgggcgccaaggacgtccacca 300
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     241 GTACTCCCAGAAGATTATACTCGTGCTGATGAAGAGCCCGGGCGNCAAGGACGTCCACCA 300

Qy     301 ggtcgccctcgtaagatgccgcgtcacgaatcttcaacttcacttatggaatcaccacgc 360
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     301 GGTCGCCCTCGTAAGATGCCGCGTCACGANTCTTCAACTTCACTTATGGAATCACCACGC 360
```

```
RESULT   4
D65944/c
LOCUS       D65944        339 bp    mRNA            EST       13-DEC-1995
DEFINITION  CELK071DZR Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk71d12 3', mRNA sequence.
ACCESSION   D65944
NID         g1116112
VERSION     D65944.1  GI:1116112
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
            Tabara,H.
  TITLE     Toward an expression map of the C.elegans genome
  JOURNAL   Unpublished (1994)
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693510.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j
            High quality sequence stop: 257.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
                     tissue_type=whole animal"
                     /db_xref="taxon:6239"
                     /map="1q"
                     /clone="yk71d12"
                     /clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT      99 a     56 c     74 g    103 t      7 others
ORIGIN
```

```
  Query Match           33.0%;  Score 314.4;  DB 24;  Length 339;
  Best Local Similarity  94.7%;  Pred. No. 2.2e-72;
  Matches  321;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy     588 atcaattcctgagctgccaattgttccatctagaatcgatgaattcgacgctacagtcga 647
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     339 ATCAATTCCTGAGCTGCCAANTGTTCCATCTAGAATCGATGAATTCGACGCTACAGTCGA 280

Qy     648 tccaagatatgaaacagatttgaaaaatgaatacattcgtcattggaaacaagtcaaaaa 707
           |||||||||||  |||||| ||  ||||||| ||||||||||||||| ||||| ||   |
Db     279 TCCAAGATATGNGACAGATNTGTGAAATGAANACATTCGTCATTGGAGACAAGNCACCNA 220

Qy     708 aggttggtgtgctcatcaacgtcgtcggactgctccccatgcaagaagcatagcattaat 767
           ||| |||||||||||| |||||| || ||||||||||||||||||||||||||||||||||
Db     219 AGGGTGGTGTGCTCATCACCGTCGTNGGGCTGCTCCCCATGCAAGAAGCATAGCATTAAT 160

Qy     768 caacaaaatctaccagcctggagagtcgaaaactgtcgagcaagcacttggtcttattta 827
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 CAACACAATCTACCAGCCTGGAGAGTCGAAAACTGTCGAGCAAGCACTTGGTCTTATTTA 100

Qy     828 aatattctaacatgtaatttcaatttatctcttactttctgatcttgctatcacatgtct 887
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 AATATTCTAACATGTAATTTCAATTTATCTCTTACTTTCTGATCTTGCTATCACATGTCT 40

Qy     888 cttatttcaaaaatctcacttaaaattcatataaataa 926
           |||||||||| ||||||||||||||||||||||||||||
Db      39 CTTATTTCAACAATCTCACTTTAAAATTCATATAAATAA 1
```

```
RESULT   5
C55654/c
LOCUS       C55654         300 bp    mRNA            EST       16-SEP-1997
DEFINITION  C55654 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
            clone yk191f6 3', mRNA sequence.
ACCESSION   C55654
NID         g2400255
VERSION     C55654.1  GI:2400255
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
            Sano,M., Miyata,A. and Nishigaki,A.
  TITLE     Expression map of the C.elegans genome
  JOURNAL   Unpublished (1996)
COMMENT     On Jan 24, 1995 this sequence version replaced gi:634347.

            Contact: Yuji Kohara
            Gene Library Lab
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 0559-75-0771
            Fax: 0559-75-6240
            Email: ykohara@ddbj.nig.ac.j.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /note="dev_stage=varied, sex=Hermaphrodite male,
```

tissue_type=whole animal"
/db_xref="taxon:6239"
/clone="yk191f6"
/clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT 85 a 52 c 64 g 99 t
ORIGIN

Query Match 31.4%; Score 300; DB 35; Length 300;
Best Local Similarity 100.0%; Pred. No. 1.3e-68;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 575 cacgtccagataaatcaattcctgagctgccaattgttccatctagaatcgatgaattcg 634
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 300 CACGTCCAGATAAATCAATTCCTGAGCTGCCAATTGTTCCATCTAGAATCGATGAATTCG 241

Qy 635 acgctacagtcgatccaagatatgaaacagatttgaaaaatgaatacattcgtcattgga 694
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 240 ACGCTACAGTCGATCCAAGATATGAAACAGATTTGAAAAATGAATACATTCGTCATTGGA 181

Qy 695 aacaagtcaaaaaaggttggtgtgctcatcaacgtcgtcggactgctccccatgcaagaa 754
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 180 AACAAGTCAAAAAAGGTTGGTGTGCTCATCAACGTCGTCGGACTGCTCCCCATGCAAGAA 121

Qy 755 gcatagcattaatcaacaaaatctaccagcctggagagtcgaaaactgtcgagcaagcac 814
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 120 GCATAGCATTAATCAACAAAATCTACCAGCCTGGAGAGTCGAAAACTGTCGAGCAAGCAC 61

Qy 815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttg 874
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 60 TTGGTCTTATTTAAATATTCTAACATGTAATTTCAATTTATCTCTTACTTTCTGATCTTG 1

RESULT 6
D72252/c
LOCUS D72252 180 bp mRNA EST 13-DEC-1995
DEFINITION CELK096B5R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA clone yk96b5 3', mRNA sequence.
ACCESSION D72252
NID g1111959
VERSION D72252.1 GI:1111959
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 180)
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On May 5, 1995 this sequence version replaced gi:798247.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@ddbj.nig.ac.j
High quality sequence stop: 362.
FEATURES Location/Qualifiers
source 1. .180
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"
/db_xref="taxon:6239"
/map="21"
/clone="yk96b5"
/clone_lib="Yuji Kohara unpublished cDNA"
BASE COUNT 60 a 20 c 37 g 60 t 3 others
ORIGIN

Query Match 18.2%; Score 173.8; DB 24; Length 180;
Best Local Similarity 97.2%; Pred. No. 1.1e-35;
Matches 175; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 748 gcaagaagcatagcattaatcaacaaaatctaccagcctggagagtcgaaaactgtcgag 807
|||||||||||||||||||||| || ||||||||||||||||||||||||||||||||||||
Db 180 GCAAGAAGCATAGCATTAATCANCACAATCTACCAGCCTGGAGAGTCGAAAACTGTCGAG 121

Qy 808 caagcacttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttct 867
|||||||||||||||| |||||||||||||||||||| |||||||||||||||||||||
Db 120 CAAGCACTTGGTCTTATNTAAATATTCTAACATGTAATNTCAATTTATCTCTTACTTTCT 61

Qy 868 gatcttgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataat 927
|||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db 60 GATCTTGCTATCACATGTCTCTTATTTCAACAATCTCACTTTAAAATTCATATAAATAAT 1

RESULT 7
AA163826
LOCUS AA163826 438 bp mRNA EST 16-FEB-1997
DEFINITION mr18g07.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:597852 5', mRNA sequence.
ACCESSION AA163826
NID g1739791
VERSION AA163826.1 GI:1739791
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 438)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:837878.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:363284
Putative full length read
vector to vector length is 453
Seq primer: -28M13 rev2 from Amersham.
FEATURES Location/Qualifiers
source 1. .438
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
/db_xref="taxon:10090"
/clone="IMAGE:597852"
/clone_lib="Soares mouse 3NbMS"
/sex="male"

```
                     /tissue_type="Spleen"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
BASE COUNT      122 a     90 c     64 g    162 t
ORIGIN


  Query Match             4.8%;  Score 45.6;  DB 29;  Length 438;
  Best Local Similarity   59.1%;  Pred. No. 0.038;
  Matches   78;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     820 cttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttgctatc 879
           ||  | || | ||| | | || | |||    ||||| |||| ||||||  ||||    ||
Db     305 CTACTGTAGAGATTGTTATATTTCATTCTTTTTTATTTCTTTCTTTCTTTTCTTTTCTTC 364

Qy     880 acatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggtttattcaa 939
           | | | | || ||  |||| ||  ||| |||||  | | | ||   |     |||| ||
Db     365 TCCTTTTTTTTTTTCAAAAATTCAAACTGTAAAAAACTTTTCAACGTTCACGATATTTAA 424

Qy     940 atacatcatctt 951
           || |    |||
Db     425 TTAAAGTTACTT 436


RESULT   8
C90616/c
LOCUS       C90616         280 bp    mRNA            EST       20-APR-1998
DEFINITION  C90616 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSI849, mRNA sequence.
ACCESSION   C90616
NID         g3060236
VERSION     C90616.1  GI:3060236
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 280)
  AUTHORS   Yoshino,R., Morio,T. and Tanaka,Y.
  TITLE     Developmental cDNA in Dictyostelium discoideum
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2152439.

            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .280
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="SSI849"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      144 a     31 c     11 g     94 t
ORIGIN


  Query Match             4.2%;  Score 40.4;  DB 39;  Length 280;
  Best Local Similarity   56.9%;  Pred. No. 0.84;
  Matches   74;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttg 874
           ||  | || ||||||  || |||  | | | |  |||   | |  || |||||  | ||
Db     253 TTTTTTTTTTTTAAACGTTTAAATTATTAATATAATGATTTTTCTAATTTCTTTTTTTT 194

Qy     875 ctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttta 934
           | |   || ||    |  | |||||    |    ||| || |||||||| |   || |
Db     193 TTTTTTAATATCAGGAAAATAAAAAAAAAAAAAACAAATTTTATATAAATATTTATTTAA 134

Qy     935 ttcaaataca 944
           || || || |
Db     133 TTAAATTATA 124


RESULT   9
C84248/c
LOCUS       C84248         466 bp    mRNA            EST       01-APR-1998
DEFINITION  C84248 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
            discoideum cDNA clone SSB385, mRNA sequence.
ACCESSION   C84248
NID         g2723854
VERSION     C84248.1  GI:2723854
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 466)
  AUTHORS   Ochiai,H., Morio,T. and Urushihara,H.
  TITLE     Developmental cDNA in Dictyostelium discoideum (46)
  JOURNAL   Unpublished (1997)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2284815.

            Contact: Hideko Urushihara
            Institute of Biological Sciences
            University of Tsukuba
            3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
            Email: d402hu@sakura.cc.tsukuba.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .466
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /map="1"
                     /clone="SSB385"
                     /clone_lib="Dictyostelium discoideum SS (H.Urushihara)"
                     /dev_stage="slug"
BASE COUNT      220 a     35 c     56 g    152 t      3 others
ORIGIN


  Query Match             4.0%;  Score 38.4;  DB 39;  Length 466;
  Best Local Similarity   54.3%;  Pred. No. 2.9;
  Matches   75;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     815 ttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatcttg 874
           ||| | |||||  | |||| | |    || ||   | |||| |   || |||   | |
Db     323 TTGTTATTATTATATTATTGTTATTATTATTTATTAATTATATNACACGTTCGATTTTAT 264

Qy     875 ctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttta 934
           |||  | ||   |||    |  |||||     ||| | |  || ||| |  ||     ||
Db     263 CTAAAAAATTGGTCTATAATTTAAAATTGGTTTTTGATAAGCANATACAATATCAACTTC 204

Qy     935 ttcaaatacatcatcttg 952
           |||||| ||   ||| |
Db     203 ATCAAATTCACTTTCTGG 186


RESULT  10
R52024
LOCUS       R52024         391 bp    mRNA            EST       18-MAY-1995
DEFINITION  yg84h10.s1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:40182 3', mRNA sequence.
ACCESSION   R52024
NID         g813926
VERSION     R52024.1  GI:813926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
```

```
            Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
            Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT
            Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 2127
            High quality sequence stops: 273 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2127   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 273.
FEATURES              Location/Qualifiers
     source           1. .391
                      /organism="Homo sapiens"
                      /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                      I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                      I - oligo(dT) primer [5'
                      AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT 3']
                      double-stranded cDNA was ligated to Hind III adaptors
                      (Pharmacia), digested with Not I and directionally cloned
                      into the Not I and Hind III sites of the Lafmid BA vector.
                      Library went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
                      /db_xref="GDB:412723"
                      /db_xref="taxon:9606"
                      /map="808D05; 9; 9q31-9q33"
                      /clone="IMAGE:40182"
                      /clone_lib="Soares infant brain 1NIB"
                      /sex="female"
                      /dev_stage="73 days post natal"
                      /lab_host="DH10B (ampicillin resistant)"
BASE COUNT      144 a     41 c     57 g    146 t      3 others
ORIGIN


  Query Match             3.9%;  Score 37.6;  DB 22;  Length 391;
  Best Local Similarity   56.5%;  Pred. No. 4.6;
  Matches   70;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy     830 tattctaacatgtaatttcaatttatctcttactttctgatcttgctatcacatgtctct 889
           | || ||||| ||||   |||||||| |    | ||  | |  || ||   | | ||   |
Db       4 TTTTTTAACAGGTAACAGCAATTTATTTAAAATTTAATTAGATTTCTTAAATAAGTAAAT 63

Qy     890 tatttcaaaaatctcactttaaaattcatataaataatgggtttattcaaatacatcatc 949
           |||| |  || || | |||| | | | | || | | |   ||   |   || |   |||
Db      64 AATTTAAGTAAACTTAGTTTATATTACTTGTATAAATTATTTTACATATTATCCTCAATC 123

Qy     950 ttga 953
           || |
Db     124 TTAA 127


RESULT  11
M80104
LOCUS       M80104         478 bp    mRNA            EST       30-JUN-1992
DEFINITION  wEST00641 Mixed stage, Stratagene (cat. #937006) Caenorhabditis
            elegans cDNA clone CEMSH48 similar to Cytochrome oxidase chain III,
            mitochondrial, mRNA sequence.
ACCESSION   M80104
NID         g272123
VERSION     M80104.1  GI:272123
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 478)
  AUTHORS   McCombie,W.R., Adams,M.D., Kelley,J.M., FitzGerald,M.G.,
            Utterback,T.R., Khan,M., Dubnick,M., Kerlavage,A.R., Venter,J. and
            Fields,C.
  TITLE     Caenorhabditis elegans expressed sequence tags identify gene
            families and potential disease gene homologues
  JOURNAL   Nature Genet. 1, 124-131 (1992)
  MEDLINE   93250983
COMMENT     Other_ESTs: wEST00642
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            Seq primer: M13 Foward.
FEATURES              Location/Qualifiers
     source           1. .478
                      /organism="Caenorhabditis elegans"
                      /note="Vector: Uni-ZAP XR; C. elegans mixed stage cDNA
                      library. Strategene catalog #937006. The library is oligo
                      dT primed and directionally cloned in the Uni-ZAP XR
                      vector."
                      /db_xref="taxon:6239"
                      /clone="CEMSH48"
                      /clone_lib="Mixed stage, Stratagene (cat. #937006)"
BASE COUNT      184 a     88 c     62 g    137 t      7 others
ORIGIN


  Query Match             3.9%;  Score 37.2;  DB 20;  Length 478;
  Best Local Similarity   53.6%;  Pred. No. 6;
  Matches   75;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy     797 aaactgtcgagcaagcacttggtcttatttaaatattctaacatgtaatttcaatttatc 856
           || ||| | |  ||    ||    ||||   | || |  | | | || ||  || | |||
Db     225 AAGCTGCCAATAAACATGTTAAAATTATACTATTAGTACATCTTTTATTTCTAAGTAATC 284

Qy     857 tcttactttctgatcttgctatcacatgtctcttatttcaaaaatctcactttaaaattc 916
           | | |  | || ||  || || | |||   ||| |   | ||| |  |   ||||| |
Db     285 TGTGATGTGCTCAAGTTACTGTAACACCACTCCTTAATAAAATANTTGTATTTAATAACG 344

Qy     917 atataaataatgggtttatt 936
           || |   ||||| || | |
Db     345 GTACACCAAATGGNTTAACT 364


RESULT  12
AA550683
LOCUS       AA550683       652 bp    DNA             EST       11-AUG-1997
DEFINITION  1920m3 gmbPfHB3.1, G. Roman Reddy Plasmodium falciparum genomic
            clone 1920m, mRNA sequence.
ACCESSION   AA550683
NID         g2320935
VERSION     AA550683.1  GI:2320935
KEYWORDS    EST.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
            Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N.,
            Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
            Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M.,
            Su,X.-Z., Thompson,J.K., Vital,F., Wellems,T.E. and Werner,E.
  TITLE     Current status of the Plasmodium falciparum genome project
  JOURNAL   Mol. Biochem. Parasitol. 79, 1-12 (1996)
  MEDLINE   97001675
COMMENT     On Apr 14, 1993 this sequence version replaced gi:716484.

            Contact: Dame JB
            Department of Pathobiology
```

University of Florida
2015 SW 16th Ave., Bldg. 1017, Gainesville, FL 32611-0880
Tel: 352 392 4700
Fax: 352 392 9704
Email: dame@icbr.ifas.ufl.edu
Seq primer: T3.

FEATURES Location/Qualifiers
source 1. .652
/organism="Plasmodium falciparum"
/note="Vector: pBlueScript SK(+); Genomic DNA, from asynchronous blood stage parasites of the cloned Honduran HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 50oC (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were polished using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBlueScript SK(+). Recombinant plasmids transformed E. coli XL1-Blue."
/db_xref="taxon:5833"
/clone="1920m"
/clone_lib="gmbPfHB3.1, G. Roman Reddy"
/lab_host="E. coli XL1-Blue"

BASE COUNT 222 a 77 c 43 g 310 t
ORIGIN

Query Match 3.9%; Score 37.2; DB 34; Length 652;
Best Local Similarity 55.4%; Pred. No. 6.1;
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy    814 cttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatctt 873
          |||  ||||  |    ||||  |||| |     |||   || |||   | |||   ||| |
Db     95 CTTTTTCTTTATATTCTATTGTAACTTCATCCTTCCTATTTTCTTCAATTTTTAAATCAT 154

Qy    874 gctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggttt 933
           ||    || ||  || |||||  | ||| | | |||||     ||||||  ||  | |
Db    155 TATAATCAATTTCAATTTTTTCACTATTCTTAATATAAAAAAATAATAAATCTTGTATAT 214

Qy    934 attcaaaatac 943
           |   ||| |
Db    215 TTCTTAATTC 224
```

RESULT 13
AI250750/c
LOCUS AI250750 342 bp mRNA EST 28-JAN-1999
DEFINITION qi36a05.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1858544 3' similar to gb:J03746 GLUTATHIONE S-TRANSFERASE, MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION AI250750
NID g3847279
VERSION AI250750.1 GI:3847279
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 342)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 976 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 279.

FEATURES Location/Qualifiers
source 1. .342
/organism="Homo sapiens"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."
/db_xref="taxon:9606"
/map="8"
/clone="IMAGE:1858544"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"

BASE COUNT 124 a 53 c 48 g 117 t
ORIGIN

Query Match 3.9%; Score 37.2; DB 43; Length 342;
Best Local Similarity 54.3%; Pred. No. 5.8;
Matches 75; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy    813 acttggtcttatttaaatattctaacatgtaatttcaatttatctcttactttctgatct 872
          || | ||  |  |||| | || |  |   |  ||  |   | |||  ||  ||  ||||
Db    179 ACATAGTAATTCTTAAGTCTTTTGTCTGATTTTTAAAGACTTTCTTATAAATTTGGATCA 120

Qy    873 tgctatcacatgtctcttatttcaaaaatctcactttaaaattcatataaataatgggtt 932
          || ||| |  |||  | |    |||| | ||||||||  |||  ||| ||  | ||
Db    119 TGTTATGATTTGTAACATTCACCAAACACCTCACTTTTGAATCTATAAAAGAATTGCACG 60

Qy    933 tattcaaatacatcatct 950
          |||   ||  |   || |
Db     59 TATGAGAAACCTATATTT 42
```

RESULT 14
C94326/c
LOCUS C94326 459 bp mRNA EST 15-JUN-1998
DEFINITION C94326 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium discoideum cDNA clone SSK839, mRNA sequence.
ACCESSION C94326
NID g3218941
VERSION C94326.1 GI:3218941
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
Eukaryota; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 459)
AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797754.

Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT - Dictyostelium discoideum cDNA project in Japan.

FEATURES Location/Qualifiers
source 1. .459
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSK839"
/clone_lib="Dictyostelium discoideum SS (H.Urushihara)"

/dev_stage="slug"
BASE COUNT 234 a 32 c 48 g 145 t
ORIGIN

Query Match 3.9%; Score 37; DB 40; Length 459;
Best Local Similarity 60.4%; Pred. No. 6.7;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 824 tttaaatattctaacatgtaatttcaatttatctcttactttctgatcttgctatcacat 883
Db 419 TTATTATTTTTTTATTTGTTATTTTTATTTATTAATTTTTTTTTTTATATTTATTTTTGTT 360

Qy 884 gtctcttatttcaaaaatctcactttaaaattcatataaat 924
Db 359 TTTTTTTATTTTTAATAATTAATTTAAAAATTTAAATATAT 319

RESULT 15
AA518097/c
LOCUS AA518097 571 bp mRNA EST 16-JUL-1997
DEFINITION vi23b01.r1 Barstead mouse irradiated colon MPLRB7 Mus musculus cDNA clone IMAGE:904585 5' similar to gb:M76124 Mus musculus EGP314 precursor mRNA, complete cds (MOUSE);, mRNA sequence.
ACCESSION AA518097
NID g2258782
VERSION AA518097.1 GI:2258782
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 571)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1395425.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:525249
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 77.

FEATURES Location/Qualifiers
source 1. .571
/organism="Mus musculus"
/strain="FVB/N"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; Tissue obtained from 8 week old mouse. Colon was harvested 72 hours after irradiation with 1400 Gys. 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. "
/db_xref="taxon:10090"
/clone="IMAGE:904585"
/clone_lib="Barstead mouse irradiated colon MPLRB7"
/dev_stage="8 weeks"
/lab_host="DH10B"

BASE COUNT 182 a 88 c 155 g 146 t
ORIGIN

Query Match 3.9%; Score 36.8; DB 34; Length 571;
Best Local Similarity 52.6%; Pred. No. 7.7;
Matches 80; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 763 ttaatcaacaaaatctaccagcctggagagtcgaaaactgtcgagcaagcacttggtctt 822
Db 228 TCAACAACCAAAAATTACCCCCCCCTAAAGGGAAAAACCTTGGGGCCACCCCTTTTTTTT 169

Qy 823 atttaaatattctaacatgtaatttcaatttatctcttactttctgatcttgctatcaca 882
Db 168 TTCCTTTTTCTTTTCTTTTTTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 109

Qy 883 tgtctcttatttcaaaaatctcactttaaaat 914
Db 108 TTTTTTTTTTTTTTTTTTTTTCACTTTAAAAT 77

Search completed: August 26, 1999, 11:03:32
Job time: 3091 sec